

University of Helsinki
Faculty of Educational Sciences
Helsinki Studies in Education 92

Tommi Wallenius

SCHOOLS, PERFORMANCE AND PUBLICITY
Contrasting the policy on publicising school performance
indicators in Finland with the other Nordic countries

Academic dissertation to be presented, with the permission of the Faculty of Educational Sciences at the University of Helsinki, for public examination in Minerva Square K226, Siltavuorenpenger 5 A, on 12 of September, 2020 at 10 o'clock.

Helsinki 2020

Pre-examiners

Professor Guri Skedsmo

Institute for Research on Professions and Professional Learning (IPP)

Schwyz University of Teacher Education, Switzerland

Department of Teacher Education and School Research (ILS)

Faculty of Educational Sciences, University of Oslo, Norway

Associate Professor Nelli Piattoeva

University of Tampere

Custos

Associate Professor Janne Varjo, University of Helsinki

Supervisors

Associate Professor Janne Varjo, University of Helsinki

D.Soc.Sci., Title of Docent Mira Kalalahti, University of Helsinki

Associate Professor Tero Erkkilä, University of Helsinki

Opponent

Professor Guri Skedsmo

Institute for Research on Professions and Professional Learning (IPP)

Schwyz University of Teacher Education, Switzerland

Department of Teacher Education and School Research (ILS)

Faculty of Educational Sciences, University of Oslo, Norway

Cover image

Tommi Wallenius

Unigrafia, Helsinki

ISBN 978-951-51-6451-3 (paperback)

ISBN 978-951-51-6452-0 (pdf)

ISSN 1798-8322

ISSN 2489-2297

Tommi Wallenius

Schools, Performance and Publicity

Contrasting the policy on publicising school performance indicators in Finland with the other Nordic countries

Abstract

In this dissertation I scrutinised the Finnish comprehensive school quality assurance and evaluation (QAE) policy with a special focus on the policy of publicising school performance indicators. The research originated from a simple notion: while no school performance indicators are publicised at the school level in Finland, in the other Nordic countries various comparable and commensurable school-specific performance indicators are publicised in the government's official web portals.

Thus, by contrasting the institutionalisation of the Finnish publicising policy with the other Nordic countries, the aim of this research was to clarify how and why Finland has been able to resist the pressures of the 'global testing culture' and the idea of publicising school-specific performance results. The following research questions were examined: 1) How are opposite publicising policies (being) justified in Finland and Sweden? (Article I); 2) How are the current publicising policies explained through historically institutionalised path-dependent elements? (Article II); and 3) How are the two core concepts that typically promote a school-specific publicising policy, accountability and transparency, manifested in the policy discourses in Finland, Sweden, Denmark and Norway? (Article III)

The research utilised the theories and concepts of various neo-institutional policy research approaches, above all, the writings on discursive institutionalism by Vivien A. Schmidt. Methodologically, all three research articles represented comparative policy research in education. In Article I, seven interviews with key policy actors in Finland collected within the *Fabricating Quality in Education (FabQ)* research project in 2007-08 were contrasted with the official policy justifications in Sweden. In Article II, the historical institutionalisation of the publicising policy in Finland and Sweden was examined through an analytical literature review. In Article III, 58 interviews with key policy actors in Finland, Sweden, Denmark and Norway collected within the *Dynamics in Basic Education Politics in the Nordic Countries (DYNO)* research project collected in 2015-17 were analysed.

Together, the three research articles showed how the deep-seated rationalities institutionalised in the policies, practices and policy discourses, strengthened further by Finland's initial PISA success, have provided a suitable platform for the Finnish policy actors to control the coordinative policy discourse on the comprehensive school QAE policy and to resist effectively the pressures to publicise school-specific performance indicators. The decline in the Finnish PISA scores since 2009 in my data did not show up as a 'critical juncture' at which new ideas started to challenge the legitimacy of the prevailing policy.

In Finland, the main policy discourse, which I have described as *the depoliticised discourse of school performance*, has continued to be effective in setting the limits for the 'appropriate' QAE policy and behaviour by controlling the concepts of accountability and transparency. It is noteworthy that in this discourse, the citizens have been guided to trust the Finnish comprehensive school system and 'prevented' from seeing themselves as eligible users of school-specific performance data.

The current comprehensive school QAE policy in Finland, for example the sample-based national level pupil testing that prevents the opportunity to draw up school rankings, has been established as taken for granted. Indisputably, the detrimental effects attached to school rankings, such as increased social segregation by naming and shaming of pupils, teachers and schools, should continue to be taken seriously. However, comparative research in the Nordic countries showed that the policy of publicising school performance indicators is more complex than the main discourse in Finland suggests. Demands concerning either governance transparency or families' equal rights to access official data may challenge the current publicising policy in future.

Keywords: quality assurance and evaluation policy, school performance indicators, governance publicity, accountability, transparency, comparative policy research in education

Tommi Wallenius

Koulujen arviointi ja julkisuus

Vertaileva tutkimus peruskoulun arviointitietojen julkistamispolitiikasta Pohjoismaissa

Tiivistelmä

Väitöstutkimuksessa tarkasteltiin kouluindikaattorien julkistamispolitiikkaa osana peruskoulun laadunvarmistus- ja arviointipolitiikkaa. Tutkimuksen lähtökohtana oli yksinkertainen havainto: Pohjoismaista vain Suomessa koulujen arviointitietoja ei julkaista koulukohtaisesti. Kaikissa muissa Pohjoismaissa arviointitietoja julkaistaan esimerkiksi kouluviranomaisen internet-sivustoilla ja samalla mahdollistetaan koulujen yhteismitallinen julkinen vertailu.

Tutkimuksen tavoitteena oli selvittää vertailemalla julkistamispolitiikan institutionalisoitumista muihin Pohjoismaihin, miten ja miksi Suomi on pystynyt torjumaan 'globaalin testauskulttuurin' paineet ja ajatuksen koulukohtaisten arviointitietojen julkistamisesta. Tutkimuksessa vertailtiin julkistamispolitiikan oikeutusta (Artikkeli I), historiallisia kehityskulkuja (Artikkeli II) sekä tilivelvollisuuden ja läpinäkyvyyden käsitteiden ilmentymistä koulutuksen arvioinnin asiantuntijoiden ja koulutuspoliitikkojen diskursseissa Suomessa ja muissa Pohjoismaissa (Artikkeli III).

Tutkimuksessa hyödynnettiin uusinstitutionaalisen politiikan tutkimuksen eri koulukuntien teorioita ja käsitteitä, mm. Vivien A. Schmidtin näkemyksiä diskursiivisesta institutionalismista. Historiallisen vertailun ohella tutkimuksen empiirinen aineisto koostui kahdesta erillisestä haastatteluaineistosta: vuosina 2007-08 FabQ-tutkimusprojektin yhteydessä kerätystä 7 asiantuntijahaastattelusta Suomessa, sekä vuosina 2015-17 DYNO-tutkimusprojektissa kerätystä 58 asiantuntijahaastattelusta Suomessa, Ruotsissa, Tanskassa ja Norjassa.

Väitöstutkimus osoitti kuinka historiallisesti vakiintuneet käytännöt ja puhutavat sekä niihin syvään juurtuneet rationaliteetit ovat yhdessä Suomen PISA-menestyksen kanssa luoneet otolliset lähtökohdat koulujen julkisen vertailun vastustamiselle. Myöskään Suomen vuodesta 2009 laskeneet PISA-tulokset eivät osoittautuneet tutkimusaineistossa taitekohdaksi, jossa arviointitietojen julkistamisen nykylinjausten oikeutusta olisi haastettu.

Tutkimuksen mukaan suomalaisten asiantuntijoiden omaksuma puhetapa, *kouluvertailut depolitisoiva diskurssi*, on kyennyt tehokkaasti kontrolloimaan koulukohtaisten arviointitietojen julkistamisen perusteluna toimivaa tilivelvollisuutta ja läpinäkyvyyttä peruskoulun arviointipolitiikassa. Olellaista

on huomioida, kuinka tämä puhetapa normatiivisesti on opastanut kansalaisia luottamaan suomalaiseen peruskouluun ja samalla onnistunut häivyttämään ajatuksen julkiseen ja vertailtavaan tietoon oikeutetuista kansalaisista.

Nykyinen peruskoulun arviointipolitiikka, kuten esimerkiksi koulujen paremmuusjärjestyslistausta estävä otospohjainen oppimistulosten arviointi, on vakiintunut suomalaisessa peruskoulupolitiikassa itsestäänselvyudeksi. Kiistatta koulujen 'ranking-listoihin' yhdistetyt haitat, kuten huoli alueellisen eriytymisen kasvusta sekä opettajien, oppilaiden ja koulujen leimautumisesta, on syytä ottaa vakavasti jatkossakin. Pohjoismainen vertailu kuitenkin osoitti, että koulujen tulosten julkisuus on Suomessa hallitsevaa diskurssia monimutkaisempi kysymys. Ylipäänsä vaatimukset hallinnon läpinäkyvyydestä tai perheiden yhtäläisestä oikeudesta saada virallista koulukohtaista vertailutietoa saattavat tulevaisuudessa johtaa kouluindikaattorien julkistamispolitiikan uudellenarvointiin.

Avainsanat: arviointipolitiikka, kouluindikaattorit, hallinnon julkisuus, tilivelvollisuus, läpinäkyvyys, vertaileva koulutuspolitiikan tutkimus

ACKNOWLEDGEMENTS

This dissertation has taken some time to reach its final version. During these years I have seen my kids grow, helped a young Afghan asylum seeker to struggle with the Finnish immigration bureaucracy, seen the Finnish national football team reach historically the Euro 2020 tournament (postponed due to the COVID-19 pandemic) and renovated our home (still not finished), among others. Most importantly, I have been privileged to meet and share these years with many wonderful people, to whom I would like to express here my gratitude – not, however, in any specific ranking order. :)

I want to thank my three supervisors, Janne Varjo, Mira Kalalahti and Tero Erkkilä for all your guidance and support within this process. Without your help I would not have been able to finalise this research. The last few months and especially the last few weeks before sending the manuscript for pre-examination were quite intense, but your encouragement and support was never-ending. Finally, the sharp comments by the pre-examiners, Professors Guri Skedsmo and Nelli Piattoeva, made me revise the text to its peak once more – thank you both.

Our Research Unit focusing on the Sociology and Politics of Education KUPOLI has been the primary academic home for my research. I want to thank you all for the great discussions and the moments we have shared in here: Janne Varjo, Mira Kalalahti, Hannu Simola, Jaakko Kauko, Sonja Kosunen, Petteri Hansen, Sara Juvonen, Anna-Maija Niemi, Heidi Vartiainen, Saija Volmari, Virpi Pakkanen, Hannele Pitkänen, Antti Paakkari, Samira Harjula, Helena Candido, Linda Maria Laaksonen, Eeva Rontu, Marja Peltola, Lauri Ojalehto, Eero Väättäin and many others.

Alongside KUPOLI I have been privileged to have a second academic home in the unit next door, in the Centre for Educational Assessment CEA, in which I was working for few years during my doctoral studies. This opportunity was extremely valuable and broadened my sociological view in the more practical work of educational assessment and evaluation. Thank you all, Mari-Pauliina Vainikainen, Jarkko Hautamäki, Sirkku Kupiainen, Risto Hotulainen, Ninja Hienonen, Raisa Ahtiainen, Meri Lintuvuori, Mikko Asikainen, Natalja Gustavson, Laura Kortesoja, Irene Rämä, Helena Thuneberg and many others.

In addition, I want to thank many other wonderful colleagues in our faculty, Aarno Kauppila, Tuuli From, Anna Rawlings, Markku Jahnukainen, Janne Sääntti, Esa Penttinen, just to name a few. From the University of Turku, Heikki Kinnari, Sonia Lempinen, Suvi Jokila, Jenni Tikkanen, Risto Rinne, Arto Jauhiainen and others, it has always been a pleasure to meet you, even in Turku.

This research has been funded by different organisations, mainly by the KASVA network and the Finnish Cultural Foundation. I am grateful for all the

financial resources. The JustED Summer schools in Oslo and Helsinki as well as the several national and international conferences have been important events at which to discuss and get feedback for furthering my research. My visit to Umeå University and the KASVA seminar held at the University of Oxford were impressive steps for a junior researcher.

In a family with two stressed PhD students and four children, managing the daily schedules can be sometimes quite challenging. I am thankful for all the help that I and Satu have received from our families and friends to run the daily business at home for Julius, Mio, Joel and Mai. My parents, Anu and Jussi, thank you for all your support you have given any time when needed and more. I know that you have occasionally wondered what do I actually do at the University – well, this text is something I have been dealing with. Thank you Satu's mother Mari and Ralf, father Risto and Eija, as well as all the other excellent babysitters, Kirsi and Kari, and Leena. A special thanks to my dear aunt Tuttu, who has not once stopped believing in my work and asked every Christmas me to report of my progress. An appropriate amount of external accountability seemed to work out here just perfect.

Of course much of the well-being takes place outside the workplace in other social networks. I thank all my friends, many from the Faculty of Social Sciences, Aleksi, Otto, Mikko, Elina, Henri, Markus, Antti, Jukka and so many others. You have many times politely tolerated my chats about my research topic, luckily now it is ready and it is time to go bowling. Thank you, Antti, Aulis and Juuso for the best possible company in all the unforgettable away-games with the Finnish football team – Oi Suomi on! Many thanks to all the members of Team SOSE, the best orienteering team in the world. I hope we have many Jukola relays ahead of us.

My biggest thanks should be aimed at Satu Koivuhovi, my beloved partner, mental coach and the mother of my children. You are the primus motor, not only in our family, but behind this research as well. I am definite that without your capability to seize the moment and push others towards the unthinkable I would never even have considered myself as a person doing a PhD. I am confident that your own dissertation will be ready very soon as well – the biggest party will only happen then.

As for a final remark, I dedicate this work to my son Joel and daughter Mai. I wish that you may grow up in a world that shares a view of a peaceful and progressive future, in which you learn to be open-minded and ready to challenge your own limits of unthinkable, whatever they may be.

In Helsinki, 3 August, 2020
Tommi J Wallenius

CONTENTS

1 INTRODUCTION.....	1
2 SCHOOL PERFORMANCE AND PUBLICITY IN THE ERA OF THE GLOBAL TESTING CULTURE	6
2.1 School performance indicators and the global testing culture	7
2.2 Publicising of school-specific performance results – a two-sided coin policy?	9
2.3 Accountability in educational governance.....	11
2.4 Transparency in educational governance.....	15
3 THE RESEARCH CONTEXT – FRAMING THE OPPOSITE PUBLICISING PRACTICES IN THE NORDIC COUNTRIES	18
3.1 Finland – Upstream QAE policy with restricted publicity	18
3.2 Publicising policies and practices in the other Nordic countries.....	20
3.2.1 Sweden.....	22
3.2.2 Denmark.....	24
3.2.3 Norway	26
4 NEO-INSTITUTIONAL POLICY RESEARCH AS THE THEORETICAL FRAMEWORK.....	29
4.1 The different schools of neo-institutionalism	29
4.2 Discursive institutionalism – From ideas to discourses.....	31
4.3 Critical juncture, ideational change and institutionalised elements	33
5 ON CONDUCTING THE RESEARCH, GATHERING DATA AND APPLYING THE METHODOLOGY	36
5.1 Evolving of the research setting.....	36
5.2 Research questions	38
5.3 On data and methodology	39
5.4 On research ethics and validity	41
6 THE RESULTS – EXPLAINING THE FINNISH PUBLICISING POLICY THROUGH HISTORICALLY AND DISCURSIVELY INSTITUTIONALISED ELEMENTS.....	45

6.1 Tracing the argumentation logic of the opposite publicising policies of school performance results in Finland and Sweden (Article I).....	45
6.2 Explaining the current national testing practices in Finland and Sweden through historical institutionalism and path-dependencies (Article II).....	48
6.3 Understanding accountability and transparency in the Nordic QAE policy discourses (Article III)	51
7 DISCUSSION AND THE KEY ARGUMENTS.....	55
8 CONCLUDING REMARKS	59
EPILOGUE.....	62
REFERENCES	64

LIST OF ORIGINAL PUBLICATIONS

This dissertation is based on the following research articles:

I Wallenius, T.J. (2015). Justifying opposite publication policies of school performance results in Finland and Sweden. In S. Jokila, J. Kallo & R. Rinne (Eds.), *Comparing Times and Spaces: Historical, theoretical and methodological approaches to comparative education* (pp. 209–231). Research in Educational Sciences 69. Helsinki: The Finnish Educational Research Association.

II Wallenius, T. (2016). Oppimistulosten kansallisen arvioinnin historiallinen institutionaalistuminen Suomessa ja Ruotsissa. [National Testing of Pupils in Finland and Sweden in Light of Historical Institutionalisation] In H. Silvennoinen, M. Kalalahti & J. Varjo (Eds.), *Koulutuksen tasa-arvon muuttuvat merkitykset. Kasvatussosiologian vuosikirja 1* (pp. 99–131). Research in Educational Sciences 73. Helsinki: The Finnish Educational Research Association.

III Wallenius, T., Juvonen, S., Hansen, P. & Varjo, J. (2018). Schools, accountability and transparency – approaching the Nordic school evaluation practices through discursive institutionalism. *Nordic Journal of Studies in Educational Policy*, 4(3), 133–143.

TABLES AND FIGURES

Tables

Table 1. Discourses and Narratives of Accountability.....13

Table 2. Use of results for accountability and development across countries.....14

Table 3. The relationship between transparency and accountability.....16

Table 4. Research setting in the original sub-studies (Article I, II and III).....38

Figures

Figure 1. Publicising of individual school results in national tests at the basic education level in 2008/09.....21

Figure 2. A screenshot of the Välja Skola [Choose the School] web portal in Sweden.....23

Figure 3. A screenshot of the Skoletal web portal in Denmark.....25

Figure 4. A screenshot of the Skoleporten web portal in Norway.....27

Figure 5. Theoretical framework in this dissertation.....33

Figure 6. Mathematics trends in the OECD’s PISA assessments in the Nordic countries.....34

Figure 7. Stephen E. Toulmin’s Model of Argumentation.....46

1 INTRODUCTION

Monitoring, auditing, regulating, quality assurance, accountability, appraisal and inspection are the new assessment mantra, the visible articulation of a brave new world in which it is assumed that the quality of social activity and institutional performance can be dissected piece by piece like a specimen upon the table, its organs laid bare to scrutiny, judgement and comparison. Likewise, the new creeds of criteria and transparency are supposed to reassure a skeptical world that the huge assessment effort now required and the associated widespread sacrifice of autonomy is justified by the evident gains they lead to in relation to both efficiency and equity. (Broadfoot 2000, xii.)

Approximately two decades have passed since these words by Patricia Broadfoot were published in the book *Assessment: Social Practice and Social Product*. In her most perceptive preface, the author described in a breath-taking manner the overwhelmed intensification of educational assessment policies and practices; the ideologies, the techniques and the promises in the quest of quality and institutional performance. Indeed, if looking at the past few decades, we have witnessed globally a significant increase of varied assessment and evaluation methods touching all educational levels from the universities to basic education (e.g. Kellaghan, Stufflebeam & Wingate 2003). This trend has inspired many scholars to describe its evolution and features as ‘the audit society’ (Power 1997), ‘performance measurement society’ (Bowerman, Raby & Humphrey 2000), ‘the global education reform movement’ (Sahlberg 2011) ‘the evaluation society’ (Dahler-Larsen 2011) or ‘the global testing culture’ (Smith 2016).

Consequently, the results of the different quality assurance and evaluation (QAE) methods are now also more visibly present in public. The universities are listed regularly according to several performance indicators on global university league tables such as the *Times Higher Education* or the Academic Ranking of World Universities, better known as the *Shanghai Ranking*. The school-specific results of the matriculation exam at the end of general upper secondary education in Finland are annually highlighted in the media. Above all, the attention paid to the OECD's (The Organisation for Economic Co-operation and Development) PISA assessments (Programme for International Student Assessment) in the 2000s, has lifted the learning results and pupils' performance globally into public awareness. Thus, in only a few decades, many educational systems around the globe have begun to operate in an environment in which different performance indicators have come to represent institutional quality.

However, not in all cases. The Finnish basic education quality assurance and evaluation culture has been characterised as an ‘upstream policy’ in the global

trend (e.g. Simola, Rinne, Varjo, Pitkänen & Kauko 2009; Simola, Varjo & Rinne 2010). The guidelines for the Finnish QAE framework were formulated during the 1990s law drafting work and manifested in the Basic Education Act in 1998. Within these guidelines, it was decided that the pupils' national testing should be conducted on a sample-based method and no school performance results should be publicised at the school level.

On the contrary, the other Nordic countries have come to practise exactly the opposite QAE policy. In Sweden, school-specific performance indicators have been publicised from the early 2000s and more recently the governments in Norway and Denmark have also started to make various school performance data publicly available. This is somewhat surprising, as the policy that enables school comparisons, ranking lists and league tables has been typically linked to the British and American educational cultures (e.g. the US, UK, also Australia) which feature high test-based accountability measures, sanctions or appraisals (e.g. Madaus, Russell & Higgins 2009; OECD 2013).

In this dissertation, I have scrutinised the institutionalisation of the Finnish comprehensive school quality assurance and evaluation (QAE) policy with a special focus on the publicising policy of school performance indicators. By contrasting the Finnish publicising policy with the other Nordic countries¹, my aim has been to clarify how and why Finland has been able to resist the pressures of the 'global testing culture' and the idea of publicising school-specific performance results. How is the legitimacy of the publicising policy constructed in Finland, in contrast to the other Nordic countries? And, theoretically, how do institutional practices sustain or change?

To deal with these research tasks, in this dissertation – including this Summary and the three original sub-studies (Article I, II and III) – I have utilised the theories and concepts developed under several neo-institutional policy research approaches (Hall & Taylor 1996), above all the writings of Vivien A. Schmidt on discursive institutionalism (Schmidt 2008; 2010; Carstensen & Schmidt 2016). The publicising policy, its practices and policy justifications are understood to have evolved in a historical process that entails institutionalised, path-dependent and established practices and rationalities. However, the prevailing policies are being challenged by new ideas and discourses, which could eventually lead to institutional change. Thus, by examining the historically, culturally and discursively institutionalised elements, the policy rationalities, discourses and solutions and their contextual premises and prerequisites, the aim with this research is to explain the Finnish 'upstream policy', a policy that is apparently different within the Nordic countries.

¹ Note: Iceland has not been included in this research, for practical reasons (see chapter 5.1).

All three sub-studies are approaches to the same general research task, yet from slightly different theoretical perspectives, using different research methods and data. In the first sub-study (Article I), I analysed the construction of the argumentation logic and the justifications of the publicising policy in Finland and Sweden. Here, seven interviews with Finnish educational experts were examined using Stephen A. Toulmin's model of argumentation (Toulmin 1958).

In the second sub-study (Article II), I investigated the development of the publicising policy in a wider historical perspective by contrasting the historical institutionalisation of the national level pupil testing in Finland and Sweden. Here, the analysis was based on a literature review of official documents and other contemporary writings.

In the third sub-study (Article III), I returned to analyse the more recent QAE discourses but in this final sub-study, the policy discourses were contrasted across four Nordic countries, Finland, Sweden, Denmark and Norway. Here, the analysis consisted of 58 interviews with key policy experts in these four Nordic countries.

The dissertation is positioned within the research area of comparative policy research in education. Contrasting Finland with the other Nordic countries has a special intention, for the following reasons. Firstly, the Nordic countries are often characterised as being relatively similar in their political, social and cultural systems. The main socio-political attributes of the Nordic model in education, a mainly public school system and emphases on social and educational equality are still from a global perspective featuring the Nordic comprehensive school systems, despite the countries' diverging responses to neoliberal policies (Esping-Andersen 1990; Antikainen 2006; Telhaug, Mediås & Aasen 2006; Blossing, Imsen & Moos 2014). Secondly, as for the publicising policy, all the Nordic countries have had a long history in supporting the idea and principles of administrative openness and governance publicity, known also as 'Nordic openness' (Erkkilä 2010; 2012). Thirdly, the other Nordic countries have traditionally been important reference countries in policymaking, whose examples and policy solutions have been watched closely (Hansen, Wallenius, Juvonen & Varjo 2019). Against these views, it is highly interesting that the Finnish QAE policy and the rationale concerning the publicising issue seems to differ significantly from the other Nordic countries. The fourth reason is more or less methodological. In my research, the other Nordic countries provide a reflective surface from which to contrast and understand the Finnish policy. By contrasting the institutionalisation of the publicising policy with the closest reference countries – Sweden, Denmark and Norway – I aim to 'go behind' the 'taken for granted', to make the Finnish QAE policy discourse and its embedded rationalities visible.

The primary empirical data in this research comprise two interview data sets with key educational experts. The educational experts – politicians, educational officials, academics and other stakeholders – are seen as the key actors in

constructing of the coordinative policy discourse (Schmidt 2008). Following ontologically the research tradition of social constructivism, the discourses of the interviewees were understood as both reflecting and shaping the construction of the social reality and the policymaking context – in other words, describing and defining what is, what ought to be and for what reason, why.

The interviews were conducted within two international research projects. In Article I, the seven interviews with Finnish educational experts were collected as part of a research project called *Fabricating Quality in European Education (FabQ)* in 2006–2009. In Article III, I analysed 58 interviews with educational experts in Finland, Sweden, Denmark and Norway with my colleagues Sara Juvonen, Petteri Hansen and Janne Varjo. These interviews were part of a later research project *Dynamics in Basic Education Politics in the Nordic Countries (DYNO)* in 2015–2018.

Analysing two interview sets in different time periods entails a special meaning for the research setting in this dissertation. This relates to the role of the OECD's PISA assessment in basic education policymaking. The Finnish school system became globally known because of its high scores in the first PISA assessments in the 2000s. However, since 2009 the Finnish PISA scores have indicated a continuous, if not yet radical decline. In many countries, low PISA scores or 'PISA shocks' have turned into demands for policy reforms, to enhance the efficiency of the school system especially by increasing accountability mechanisms and intensifying pupils testing. Thus, one important motive for this research was to evaluate whether this decline would mark a critical juncture or turning point, in which new ideas and discourses arise to challenge the prevailing policies and practices.

However, together the three research articles showed that the deep-seated rationalities institutionalised in the policies, practices and policy discourses, strengthened further by Finland's initial PISA success, have provided a suitable platform for the Finnish policy actors to control the coordinative policy discourse on the comprehensive school QAE policy and to resist effectively the pressures for publicising school-specific performance indicators. The decline in the Finnish PISA scores in my data did not show as a critical juncture to challenge the legitimacy of the prevailing policy. On the contrary, the dissertation shows how the widely adopted discourse that I have described in this dissertation as *the depoliticisive discourse of school performance* manages effectively to fade the need for policy reform by controlling the discursive manifestations of the two central concepts, accountability and transparency, through which the school-specific publicising policy is typically promoted. What may be most important thing to note here for policy legitimation is that within this discourse, policy development has been such that the citizens (parents) have been guided into thinking of themselves as not being eligible users of school-specific performance

data. This notion, as discussed in the concluding chapter, may however become problematic in the future despite its normative status in current society.

In this research, I have consciously taken a critical stand towards the Finnish QAE policy, the policy discourse and its embedded rationalities. To inform the reader, a critical approach does not automatically imply criticism, yet I fully understand the potential of misreading this – the topic is highly delicate and emotive, especially within the educational sector. My approach has not been the most conventional in that sense, especially as the Finnish comprehensive school system has become a source of national pride. Thus, I want to underline that my aim in this research is not to take a normative stand and to evaluate what policy ought to be practised, but rather to go behind the ‘taken for granted’, to scrutinise and understand the elements and mechanisms through which the current publicising policy and its practices have institutionalised and are being legitimated in Finnish society.

Positioned within the research area of comparative policy research in education, this dissertation consists of three peer-reviewed research articles and this Summary. For this Summary, I have extended and clarified the theoretical framework and the conceptual definitions, which in my mind have been discussed too narrow in the original research articles. The content of this Summary is as follows: In Chapter 2, I present the key concepts within this research. Chapter 3 presents my research context and the current publicising practices in the Nordic countries. In Chapter 4, I illustrate the theoretical framework of my research. In Chapter 5, I present my research journey, the empirical data and the choices concerning research methodology. Chapter 5 ends with a reflection on research ethics. In Chapter 6, I present a short summary and the main findings of each original research article. Finally, in Chapter 7, I present and discuss the key arguments of my dissertation and Chapter 8 ends the dissertation with some concluding remarks. The original research articles are attached after this Summary.

2 SCHOOL PERFORMANCE AND PUBLICITY IN THE ERA OF THE GLOBAL TESTING CULTURE

In everyday language, the publicising policy² of school-specific performance results is often associated with school rankings. Undoubtedly, the school rankings in the media are the most visible affirmation of this policy. However, in my thinking this view is too narrow, especially as in recent years, governments have taken a more active role in data publicising. For this research, I defined the publicising policy in a more broader sense:

As a constitutive element of the quality assurance and evaluation (QAE) policy, the publicising policy constructs the interface, in which the school system as a closed institution becomes under broader surveillance and in which a single school may become measurable and commensurately comparable with other units by its performance indicators in society.

This definition entails three interconnected tiers: First, the policy defines *who* of all the actors in society are thought to be entitled to access the information. Secondly, the publicising practices define *how* the data are made accessible. And thirdly, the policy justifications express *how and why* the publicising policy is legitimated, reflecting to the underlying institutional rules and rationalities in society. All these tiers are present in my research, even if my main interest is in the last tier.

In this second chapter, I hope to lead the reader to perceive an overview of my research topic and its core concepts. To start, in 2.1, I have discussed the growing use of performance indicators in education and the emergence of the so-called ‘global testing culture’ (Smith 2016), its premises and features. After that in 2.2, I have continued to discuss more closely the problematic nature of the publicising policy as a ‘two-sided coin policy’, entailing both supporting and resisting arguments and policy justifications. In the last few subchapters, I have scrutinised the concepts of accountability (2.3) and transparency (2.4) more closely, both of which have become central in recent educational governance, being also the core concepts by which the need to publicise school-specific performance data is now promoted. I have explained how these two concepts in educational governance are

² Throughout this Summary, I have chosen to use the verb ‘to publicise’ [to bring to the attention of the public] instead of ‘to publish’ [to disseminate to the public] to illustrate the political nature of school performance data. Even though the performance data can be understood as an official information *per se*, presenting single schools in a comparable form entails a nature of an intended policy to draw attention in their performance.

closely connected and come together in the concept of institutional ‘answerability’ (Fox 2007).

2.1 School performance indicators and the global testing culture

Quality has come to play a central part in all educational systems. Or, as Kauko, Rinne and Takala (2018) put it, ‘quality assurance and evaluation’ (QAE) has continuously become more important in educational policy and governance around the globe in recent decades. The range of evaluation practices through which the economy, effectiveness and efficiency of educational systems are monitored, is wide. These include school inspections, international large-scale assessments, national and local level pupils’ assessments and schools’ self-evaluations, among others. During the last 30 years, we have witnessed enormous growth in the use of performance indicators in all public sector governance, including education. One central feature of this evolution has been the growing use of school performance indicators (SPI), by which individual units, the schools, can be observed and compared in a commensurable form.

In its widest sense, ‘a performance indicator’ is a summary statistical measurement on an institution or system, which is intended to be related to the quality of its functioning (Goldstein & Spiegelhalter 1996). Such measures may concern several aspects of the system and reflect a range of objectives. ‘Input’ indicators such as the pupil/teacher ratio are often used to estimate the resources available to institutions. ‘Process’ measures such as average teaching time per pupil may reflect the organisational structure, whereas ‘outcome’ measures such as school examination results have been used to judge institutional ‘effectiveness’ (Goldstein & Spiegelhalter 1996).

In the educational context, pupils’ learning achievements are the clearest example of a statistical performance indicator. Pupils’ learning achievements are constantly monitored through a range of assessments at the local, national and international levels. QAE has evolved into a fourth mechanism for steering educational systems, along with the legal, economic and the ideological systems around the globe (Lundgren 1990).

The technological development of statistical analysis software and internet web portals has made the management and illustration of statistical data more convenient during the past few decades. However, the main reason behind the development of performance indicators has expanded from the governments’ continuous need to monitor and improve the quality, results and efficiency of the institutions and their activities. The expansion in the use of pupils’ standardised testing and other statistical measures of school performance in educational governance has led many researchers to describe the trend in a critical tone either as ‘the audit society’ (Power 1997), ‘the performance measurement society’ (Bowerman, Raby & Humphrey 2000), ‘the global education reform movement’

(Sahlberg 2011), ‘the evaluation society’ (Dahler-Larsen 2012) or ‘the global testing culture’ (Smith 2016).

According to Smith (2016), the emergence of a global testing culture derives from two core assumptions: ‘positivism’, which assumes that the reality can be observed and objectively measured, and ‘individualism’, which emphasises the idea of a rational and independent actor who is capable of making successful manoeuvres if given the necessary information. These two underlying core assumptions form the basis for the cultural values (‘education as a human right’, ‘academic intelligence’, ‘faith in science’, ‘decentralisation’ and ‘neoliberalism’), through which the need of the policy practices, e.g. standardised testing, accountability and results publication, are promoted. The values eventually outline the legitimate behavioural guidelines of all the actors associated with education, foremost the pupils, teachers, parents and the government. As Smith put it,

Characterised by census-based standardised testing with links to high-stakes outcomes, the global testing culture can be seen in the expansion of testing and accountability systems around the world and the increasingly ‘common-sense’ notion that testing is synonymous with accountability, which is synonymous with education quality. (Smith 2016, 7.)

The expansion of the various quality assurance activities has been generated around a global educational policy discourse in which evaluation is seen as a central tool for governments to monitor and to improve the quality of the school system and thus the countries’ economic competitiveness in the globalised world (Rizvi & Lingard 2010). This view has evolved alongside other market logic driven ideologies, often described with concepts such as neo-liberalism, new public management or quasi-markets in education. The spread of the evaluation policies and practices has been explained by the processes of ‘policy borrowing and lending’ (Steiner-Khamsi 2004), ‘policy learning’ (Lingard 2010) and ‘travelling policies’ (Ozga & Jones 2006), to name a few. Many researchers have pointed to the role of the international organisations (e.g. the OECD, European Union, IMF, World Bank) as influential actors in defining and mediating the policy of ‘best practices’ in the field of QAE. Especially the OECD has explicitly promoted the importance of a systematic and holistic evaluation framework to monitor the quality of educational systems (e.g. OECD 2013; see Niukko 2006).

However, despite the global policy flows or trends, education is a field built on relatively strong domestic policymaking. Even though we may accept the notion of a relative convergence in the QAE policy and practices at the global level, the differences between countries’ QAE policies and practices are significant. The globally travelling ideas and policies tend to reshape, as they meet the ‘deep-seated historical traditions institutionalised in the structures, practices

and institutional cultures [that] are specific to each nation' (Green 1997, 23; see also Simola et al. 2009). This notion is at the very heart of my research approach and is also shown in all the three original sub-studies.

2.2 Publicising of school-specific performance results – a two-sided coin policy?

Pupil assessment through standardised testing has a long tradition in educational governance, dating from the early 20th century (Eurydice 2009; Shiel, Kellaghan & Moran 2010). However, publicising pupils' assessment results and other school performance data in a comparable manner is a relatively new phenomenon in the educational sector. In Europe, the policy was first introduced in England in the 1980s as one of many market-oriented educational reforms made by the Conservative governments. Schools were first legally obliged to publish details of their examination results under the Government's School Information Regulations in 1981 but it was not until 1991 that the results were required to be published in 'a common and consistent form' (Department of Education and Science, 1991). In the following year (1992), the results for individual schools were first produced by the Department for Education (DfE) and publicised in the national press. Finally, in 1993, the results of independent (private) schools were publicised alongside those of state-maintained schools (West & Pennell 2000; Hallgarten 2001).

Despite the criticism, the new policy of national league tables was seen as being needed. The school ranking lists formed a key strand of the parents' rights defined in the Parents' Charter (Hallgarten 2001; Beveridge 1992). The foremost and explicit intention was to keep the parents informed and to serve the parents in their school choice:

Under the Government's reforms you should get all the information you need to keep track of your child's progress, to find out how the school is being run, and to compare all local schools. (Department for Education 1994, 3.)

The school performance tables (ranking lists, league tables) in the media are the most public affirmation of the trend of global testing culture. Their use has been characteristic especially in the British and American educational cultures (e.g. the US, UK, Canada, Australia) featured with neo-liberal educational policy, marketisation, consumerism, choice, competition and high test-based accountability measures, sanction or appraisal procedures (e.g. Madaus, Russell & Higgins 2009; OECD 2013). However, during the 1990s and the 2000s, more countries have adopted similar QAE policies and practices, but with national characteristics (Eurydice 2009). The questions of how and for whom the results ought to be publicised, has become a heavily debated policy issue in many

countries. Thus, in the literature, the publicising policy has been characterised as a ‘two-sided coin policy’ entailing both supporting and resisting policy arguments (Karsten, Visscher & De Jong 2001; Visscher 2001; Rosenkvist 2010, 30).

The need to publicise school-specific evaluation data has been argued primarily according to two ideas or motives: firstly, to provide relevant information on individual schools for the parents to support their school choice, and secondly, to increase accountability in the school system by drawing wider attention in the society in school performance. Here, relevant information refers to official (statistical) data, in other words ‘cold’ formal knowledge against the unofficial or ‘hot’ knowledge (hearsay, experiences of friends etc.) in parents’ school choice (Ball & Vincent 1998).

In contrast, in the view of those resisting this, publishing school-specific data has been criticised mainly on three grounds. The ‘technical-analytical critique’ has pointed out several shortcomings in the validity of evaluation information as well as in the publishing practices of the results (Visscher 2001, 202-204). This type of critique has led to the development of the so called value-added measurements, the aim of which is to control several of the pupil or school background variables (e.g. socio-economic status, parents’ educational level, average household income of the school district etc.) and to provide more accurate information on schools’ relative performance. However, despite many technical improvements, the value-added indicators have not removed the criticism of other types. The critique focusing on the ‘usability problems’ has pointed out that the accessibility of school performance publications is not equally distributed across all parents and the complex data are often difficult to interpret (Visscher 2001, 204-205).

However, the loudest criticism has focused on the effects of school-specific publication, especially school rankings. This type of ‘political-ethical and societal critique’ has warned that paying excessive attention to school performance may lead to unintended consequences, for example to a narrow ‘teaching to the test’ view within school work or to questionable changes in the selection criteria in pupils’ intake. Above all, the ethical and societal critique has questioned the validity of the market-driven competition logic, meaning here the idea of a causal relationship between school-specific publications and the improvement of the schools’ performance. On the contrary, it has been argued that the publicising policy (in everyday language, the school ranking policy) has accelerated the social segregation of schools and their neighbourhoods. The ‘naming and shaming’ effect is often used to describe a process in which different actors, e.g. parents or teachers start to avoid the schools with lower performance results (Visscher 2001, 205-6; see also Karsten et al. 2001; Maw 1999; van Petegem, Vanhoof, Daems & Mahieu 2005; Power & Frandji 2009; Simola 2005; West & Pennell 2000.)

As Hallgarten (2001) has noted, the actual effects of the school ranking policy, either within the schools or more broadly in society, are difficult to extrapolate

from the other simultaneous policy reforms that have been characteristic of the neo-liberal educational policy of recent decades (e.g. school choice, school profiles, increased accountability etc.). Research on the effects is somewhat centred on the US and UK contexts, thus the findings are more or less context related. Some research has shown that increased accountability has positive impacts on American pupil achievement (e.g. Hanushek & Raymond 2005) (be it with reservations), or that in Wales, the abolition of the school ranking list in 2001 reduced the average pupil performance and even increased educational inequality (Burgess, Wilson & Worth 2010). Yet, it is reasonable to accept the notion that the tangible nature of competition between schools and intensified accountability measures increases through the publicising policy by putting the schools, teachers and pupils under constant performance pressure (Elstad 2009; Perryman, Ball, Maguire & Brown 2011). Thus, it is no wonder that many educational and sociological academics have taken a critical view concerning school rankings, growing emphasis on school performance indicators and intensified pupil testing policies (e.g. Rizvi & Lingard 2010; Simola 2005).

However, despite of all the criticism, the attention paid to school performance indicators and pupils' learning results has spread since the 1990s continuously. To understand better the implementation of this controversial policy, it is important to note how the publicising policy relates to the two powerful ideas and discourses in the current (educational) governance: *accountability* and *transparency*.

2.3 Accountability in educational governance

Accountability has become a cornerstone in the public sector reforms of many countries, including governance in education (e.g. Rosenqvist 2010). Originally, the concept of accountability had its roots in the sphere of financial accounting but was adopted in wider use with the new public management (NPM) reforms of the 1980s. As the context and the use of the term has expanded, its coverage has also become more varied and mixed (see Mulgan 2000; Dubnick 2014; Sinclair 1995; Erkkilä 2010).

In general, the underlying rationale in accountability is that the producers should be held accountable for the outcomes they generate. In its core sense, accountability can be defined as a process of 'being called to account to some authority for one's actions' (Jones 1992, 73). This definition of accountability has several features: it is external – accountability takes place between the external account holder, an actor or constituency holding someone such as an official or institution, accountable; it involves social interaction and exchange – the actor seeks answers and rectification while the other side is opposed to responds and sanctions; and it implies rights of authority – the actor that is calling for the accounting has authority over the one who is to be held accountable (Mulgan 2000, 555).

The actors and stakeholders related to QAE activities can be understood as forming a hierarchical chain of accountability layers ranging from below (Bracci 2009). In the context of school performance indicators, the pupils' achievements, e.g. the standardised national test results, are produced in a class led by a teacher in a given school run by a principal. In a decentralised system, the evaluation results may be used by the local authorities but because of the idea of reciprocity between autonomy (local) and accountability (state), the national level actors (ministry and policymakers) may be seen as the primary users of the test results. However, the highest level of the hierarchy can be thought of as representing society, as in a democracy the elected political representatives, decision-makers and officials can be seen as being accountable or answerable by serving and informing the society of their core functions.

The idea of accountability as a tool of governance has come to permeate both the public and the private sectors. In general, accountability has been called for to increase the actors' (or institutions') efficiency and responsibility to society. According to Dubnick (2014), the generalised use of accountability as a cultural keyword lies exactly in its chameleon-like nature (Sinclair 1995), in its capability to adapt different policymaking contexts and varying policy discourses³.

The different but overlapping variations of accountability can be shown in Table 1 (Dubnick 2014). Most important thing to notice is that each discourse has its own special narrative, a promise of 'something better'. In the context of educational governance and school evaluation, we may think that at least three of these four variations – incentivisation, mechanisation and institutionalisation – shape and construct the discursive practices. Accountability in the form of standards and metrics (e.g. school performance indicators) entails a promise of enhancing the performance of the school system; accountability as administrative means and rules (e.g. QAE activities in whole) is promised as a means to support the control over the decentralised school system; while accountability as arrangements has a promise of democracy, referring here to a broader view of answerability between the school evaluation institutions and society. Moreover, as Dubnick (2014, 8) argues, the nature of these discourses and the 'promising narratives' is, that they are not only descriptive, but also generate the need for reform:

Whether focused on attaining higher ends (democracy and justice) or basic means (facilitating control or enhanced performance), discursive forms of accountability are closely tied to efforts to bring about change and reflect views that

³ The connotation of efficiency has touched exactly on the public sector, whereas the demands to increase responsibility has been directed rather to the private sector. Depending on the political issue, the need for accountability has been expressed equally by both the political left and the political right.

accountability (however it is defined) is either lacking or insufficient under current circumstances. (Dubnick 2014, 8.)

Table 1. Discourses and Narratives of Accountability (Dubnick 2014)

Discourse focused on	Narrative	Accountability as	Examples
Institutionalisation	Promise of democracy	Arrangements (usually constitutional) intended to constrain power and foster answerability and responsiveness of officials.	Constitution making; Self-restraining State; Accountability forums; Horizontal accountability
Mechanisation	Promise of control	Means used to oversee and direct operations and behaviour within organised context	Administrative control; Bureaucratisation; Rules; Reporting; Auditing
Juridicisation	Promise of justice	Formalisation (usually legal in nature) of rules and procedures designed to deal with undesirable and unacceptable behaviour.	Criminalisation; Enforcement; Truth & Reconciliation
Incentivisation	Promise of performance	Standards and metrics designed to influence behaviour.	TQM; Performance measurement; Performance management; Standards

However, in what manner the concept of accountability is perceived – its usefulness, benefits or disadvantages, is highly context related. Even within the educational sector, we may have a range of perceptions of the mechanisms and usefulness of accountability at different levels of the educational system. Especially in basic education, the idea of accountability is problematic and controversial, for several reasons.

First, we may justifiably ask how pupils' learning results can in principle be treated as a performance indicator, to explain the quality of an individual school. Furthermore, linking the pupils' assessment results with high accountability measures, sanctions or appraisals such as school closures or teachers' bonus salaries, can be seen as being extremely questionable.

By its definition, the concept of accountability always entails a certain element of power, control or reciprocity. However, countries differ in their aims and

practices set for QAE policy, as well as in the level of accountability functions. One way to approach the use of evaluation results is to estimate the level of accountability and its balance with developmental functions (Table 2).

Table 2. Use of results for accountability and development across countries (OECD 2013, 64)

Use of results for accountability	Use of results for development		
	High	High	Moderate
		Australia, Chile	Mexico, Slovak Republic, Sweden
		Belgium (Fl.), Canada, Israel, Korea, New Zealand	Czech Republic, France, Hungary, Ireland, Netherlands, Poland, Portugal
	Low	Denmark, Iceland, Norway	Austria, Belgium (Fr.), Estonia, Finland , Italy, Luxembourg, Slovenia, Spain

Countries with a strong focus on accountability typically emphasise high-stakes standardised assessment of students; teacher appraisal that is linked to decisions regarding career advancement, salary, promotion and dismissal; external reviews or inspections of school quality; and publication of school evaluation results and/or public comparisons of school performance. On the other hand, countries with a strong focus on development and improvement typically emphasise formative, low-stakes assessment of students; teacher appraisal that is linked to decisions regarding teacher professional development and learning opportunities; and school self-evaluation and external support for organisational learning (OECD 2013, 64).

Table 2 shows how the Nordic countries are positioned in a two-dimensional typology based on the use of evaluation and assessment results. Sweden is characterised as ‘high for accountability’ but ‘moderate for development’. Denmark, Iceland and Norway are all grouped as ‘low for accountability’ but ‘high for development’ whereas Finland is categorised in the low-right corner as ‘low for accountability’ and ‘moderate for development’.

It is important to note the categorisation of levels of accountability functions within this research. How do the policy actors understand the concept of accountability in the basic education QAE policy in the Nordic countries? I will discuss this mainly in Article III (see Chapter 6.3). Meanwhile, the following subchapter turns to a discussion of the other essential concept, through which the publicising policy is being advocated, namely transparency.

2.4 Transparency in educational governance

Along with accountability, the concept of transparency has become at least as important in the current politics and governance. In a wider historical perspective, the idea of institutional and administrative ‘publicity’ or ‘openness’, has been central in the evolution of the modern state and democracies but in recent decades, the use of ‘transparency’ has expanded in the policy discourses and shifted to resonate more with the ideas and demands for institutional efficiency (Erkkilä 2010)⁴.

Similar to the concept of accountability, the definition and the meaning of transparency is also multifaceted and ambiguous (Ball 2009; Hood 2010; Bauhr & Grimes 2012). Monika Bauhr and Marcia Grimes capture the idea of good governance well in their definition, as they see:

transparency as the availability of, and feasibility for actors both internal and external to state operations to access and disseminate information relevant to evaluating institutions, both in terms of rules, operations as well as outcomes (Bauhr & Grimes 2012, 5).

Both terms, accountability and transparency, are closely linked to each other. In general, transparency is often thought to precede and enhance institutional accountability. However, their further connection with improved institutional performance or increased institutional trust in society has been questioned in many pieces of research (e.g. Grimmelihijsen, Porumbescu, Hong, & Im 2013).

As Jonathan Fox (Fox 2007) has noted, the question is rather what kind of transparency will lead to what kind of accountability, and, under what conditions? Table 3 illustrates the relationship between transparency and accountability. For Fox (2007), opaque transparency refers to ‘raw data’ that is often difficult to understand, whereas clear transparency refers to information that has been made more comprehensible and accessible to the public to tell on institutional behaviour and performance. Yet, this type of clear information does not automatically need to be linked with the high level of accountability measures discussed earlier. Instead, transparency and accountability come together in the idea of institutional ‘answerability’, combining the idea of a transparent governance and soft accountability, without any direct sanctions, compensations or remediations.

⁴ According to Carolyn Ball (2009), transparency started to occur more commonly as a term for governance during the 1990s in the documents of supranational organisations (the European Union, GATT etc.) and nongovernmental organisations (NGO). Finally, the decision to name the new global civil society organisation in 1993 to investigate acts of corruption as Transparency International fostered the use of the term globally. Since then, the term transparency has gradually been adopted in general use and established its meaning to resemble the terms ‘openness’ or ‘good governance’ (Ball 2009).

Table 3. The relationship between transparency and accountability (Fox 2007, 669)

Transparency		Accountability	
Opaque	Clear	Soft	Hard
Dissemination and access to information			
	Institutional ‘answerability’		
			Sanctions, compensation and/or remediation

In the context of QAE policy and governance in basic education, the idea of transparency is promoted in at least three overlapping forms or promises. First, in its widest sense, information on the schools’ performance can be seen as a fundamental principle to enhance the citizens’ democratic rights to access information. A second view, which is linked closer to accountability, promotes the idea of institutional answerability. In this view, the citizens are entitled to information, which tells of the performance of a publicly, tax-based funded school system and its units. Here, the citizens are understood to be active members of society, entitled to observe its main functions. Thirdly, transparency supports the idea of parents’ school choice. In order to promote equal opportunities for choosing the school, the parents must be given equal access to relevant information on different schools.

Within the context of this research on Finland and the Nordic countries, governance publicity (or transparency) has a special meaning. Historically, the Nordic countries have been global forerunners in terms of administrative publicity. In general, access to government documents in the Nordic countries makes a constitutional principle of governing, often called under terms ‘principle of publicity’ or the ‘Nordic openness’ (see Erkkilä 2010; 2012). Against this background, it is highly interesting that the Finnish policy of publicising school performance indicators differs substantially from the other Nordic countries. This notion leads one to think about whether there have been certain historically embedded institutional formal and informal rules, established practices, discourses and rationalities that distinguish Finland from the other Nordic countries in terms of the basic education QAE culture and more generally concerning the basis of public information or the relationship between the state and civil society.

To summarise so far, both the concepts accountability and transparency have become central tools of governance to promote the publicising of school-specific performance indicators. Discursively, their power relies greatly on their different forms of promises, whether associated with democracy, access to information, institutional efficiency or parental choice. Furthermore, both concepts are not only

descriptive, but they also generate the need for reform. Demands for either intensified accountability measures or for increased public information are typically expressed if the institutional performance does not meet the expected performance. Thus, their use as a tool of governance is often relational to the image of the quality of the given institution, in other words 'the need to fix'.

Before moving on to present a closer description of my research, the next contextual chapter will present how the Nordic countries can be seen in their current publicising policies and practices in the comprehensive school QAE.

3 THE RESEARCH CONTEXT – FRAMING THE OPPOSITE PUBLICISING PRACTICES IN THE NORDIC COUNTRIES

The Nordic countries are often characterised as being relatively similar as for their political, cultural and societal functions. The features of the Nordic comprehensive school systems – a strong emphasis on equality, no educational dead-ends and a free of charge basic education – are often described as one of the the key components of the ‘Nordic model’ societies (e.g. Antikainen 2006; Blossing, Imsen & Moos 2014; Esping-Andersen 1990; Telhaug, Mediås & Aasen 2006). However, there are many national differences between the Nordic countries, such as in their diverging responses to neoliberal policy developments or in the aims and organisation of national level QAE policy and governance (e.g. Eurydice 2009; Hudson 2007; Ozga, Dahler-Larsen, Segerholm & Simola 2011). These differences in the educational governance may be detected in various forms, for example in the use of school inspections, standardised pupil testing – or in the use of the QAE results, that is the publicising policies and practices.

In this Chapter 3, I have provided a contextual overview of the current QAE publicising practices in the Nordic countries. The purpose of this chapter is to illustrate to the reader the different publicising practices and the implementation of the official web portals for data management within the Nordic region.

3.1 Finland – Upstream QAE policy with restricted publicity

The Finnish comprehensive school system has become known worldwide for its high-level performance in the OECD’s PISA assessments. Numerous delegations have visited Finland to examine the Finnish schools in order to elicit the factors behind the success. The Finnish QAE culture with no standardised pupil testing and no school rankings has been highlighted as one important explanatory factor (e.g. Sahlberg 2011).

The framework of the current QAE policy and practices was designed in Finland during the 1990s. Even though the role of evaluation in educational governance was strengthened in general, the Finnish evaluation practices were eventually implemented in the spirit of low accountability and control mechanisms. Despite the heavy decentralisation in Finnish governance, strong evaluation activities at the national level were left aside. The school inspection system was abolished, and it was decided that the national level pupil testing

would be conducted by a sample-based testing system⁵, leaving more space for evaluation activities at the local level in the municipalities (Varjo, Simola & Rinne 2016; Pitkänen 2019).

The principle of not publicising any commensurable and comparable performance data of individual schools, took a decisive role as the guidelines for the Finnish QAE policy were formulated during the 1990s (Jakku-Sihvonen 2013). The avoidance of school rankings by a sample-based testing system was a central policy aim, outlined first in the reports of the National Board of Education (NBoE 1995; NBoE 1998a; NBoE 1998b) and manifested finally in the Basic Education Act in 1999, articulating that *‘The main results of evaluations shall be published’* (Law 628/1998, §21).

The Finnish comprehensive school QAE policy and culture has been widely studied from a range of theoretical perspectives and with various focuses in the academic literature: Laukkanen 1997; Konttinen 1995; Räisänen 2013; Varjo, Simola & Rinne 2016; Simola, Rinne, Varjo, Pitkänen & Kauko 2009; Aurén & Joshi 2016; Pitkänen 2019, to name just a few. If there is one common view that is widely shared, it is a notion that the Finnish school evaluation culture is notably more moderate than in many other countries. The Finnish school evaluation culture has been called as an ‘upstream policy’ against the global trend of intensified evaluation practices (Simola, Varjo & Rinne 2010; Sahlberg 2011) and in many official texts (e.g. legislation, government’s documents), its function is repeatedly defined as ‘evaluation for developmental purposes’ (Varjo, Simola & Rinne 2016) – referring practically to its features of low accountability and control mechanisms.

Simola, Rinne, Varjo, Pitkänen and Kauko (2009) traced the evolution of ‘the Finnish model of QAE’ between the transnational policy pressures and national (and local) policymaking. The authors argued that the development of the Finnish policy and its practices during the 1990s was a combination of conscious, unintended and contingent factors in the changing environment of basic education governance. To sum up, the authors argued that at the national level, the Finnish QAE discourse has at least four specific characteristics:

- 1) the evaluation is ‘for developing educational services and not an instrument of administrative control’.*
- 2) the information produced through evaluation serves the administrative bodies and the schools rather than the public or families.*
- 3) practically no education official or politician has supported the provision of ranking lists or making schools transparent in competition by comparing them in terms of average performance indicators.*

⁵ The representative sample for national testing is taken regionally, covering approximately 5 to 10 percent of the age cohort and 15 percent of the schools, that is 4000-6000 pupils (Ouakrim-Soivio 2013, 20).

4) *Finland has not followed the Anglo-Saxon accountability movement in education, which advocates making schools and teachers accountable for learning results.* (Simola, Rinne, Varjo, Pitkänen & Kauko 2009, 171–173.)

These four characteristics have enjoyed of a strong consensual acceptance in Finnish society and the alternative discourses so far have clearly been marginal. However, as the authors note, the legitimacy of these principles is being tested by exogenous pressures:

until now, the Finnish antipathy towards ranking, combined with a bureaucratic tradition and a developmental approach to QAE strengthened by radical municipal autonomy, have represented two national and local embedded policies that have been rather effective in resisting a trans-national policy of testing and ranking. It is significant, however, that both of these are curious combinations of conscious, unintended and contingent factors. Therefore it also seems evident that the articulated unity these practices constitute is rather fragile given the exogenous trends and paradigm convergence of the GERM⁶. (Simola et al 2009, 174.)

The abovementioned interpretation and the four characteristics of the Finnish QAE guidelines provided a reasonable origin for this dissertation. However, in my research I continued to investigate the elements through which the legitimacy of these policy guidelines is either supported or challenged as the environment for policymaking changes. Before moving on to the theoretical framework of my research in Chapter 4, I now turn to show how the school performance indicators are publicised in the other Nordic countries.

3.2 Publicising policies and practices in the other Nordic countries

In 2009, the Eurydice network⁷ published a comprehensive report on national testing policies and practices in the European countries (Eurydice 2009). In general, the report shows how varied and mixed the European countries are as for their aims and practices set for national level pupil testing. Figure 1 shows, that in most countries, the pupil testing results are not publicised at the school level. Yet, if looking at the Nordic countries, we may notice that in Sweden, Denmark and Iceland, the results of the national tests are required to be publicised school-specifically by the central (or the local) government. In addition, since the publication of the report in 2009, Norway has also started to publicise national

⁶ For GERM (Global Education Reform Movement), see e.g. Sahlberg (2011).

⁷ See <https://eacea.ec.europa.eu/national-policies/eurydice/>

testing results at the school level (see Chapter 3.2.3). Thus, of the five Nordic countries only Finland does not publicise data on individual schools.

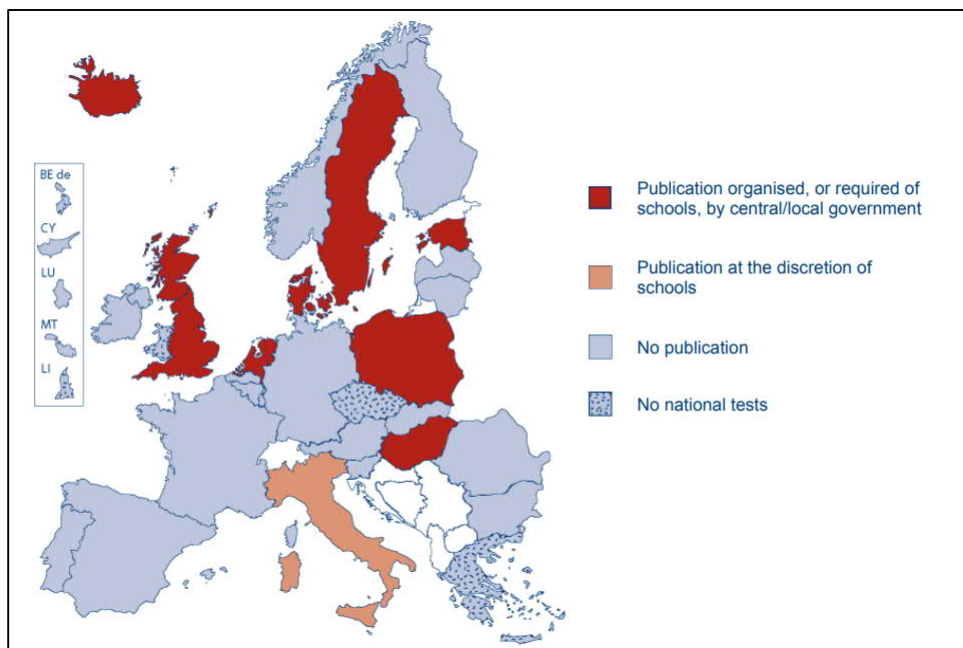


Figure 1. Publicising of individual school results in national tests at the basic education level in 2008/09 (Eurydice 2009, 54)

The policy about publicising the national tests is naturally dependant on the national testing method itself. As mentioned before, the Finnish decision to run sample-based testing has meant in practice that no such data are even collected at the basic education level, in which the learning results of all the schools could be publicised and compared commensurately⁸. The only standardised test in the Finnish education system is the matriculation exam at the end of the upper secondary school, the results of which are widely reported each year in the media.

However, as the following short overviews on Sweden, Denmark and Norway show, the publicising does not only concern the national level pupil testing results, but also various other school performance indicators. The overviews show how the governments have taken an active role in the 2000s in defining the publicising practices by launching official web portals to present the data to the public. This evolution has led to a somewhat paradoxical outcome. Despite the official instructions have repeatedly reminded everyone that the indicators ought not to be used to create school rankings, the opportunity to sort the data into ranking orders

⁸ The sample-based pupil testing is a relative rare practice among the European countries. Based on the Eurydice report (Eurydice 2009, 27), in 2009 only Finland, Spain, Latvia and Austria conducted the national level pupil testing by a sample-based testing method.

has frequently been used by the media or other third parties in all three countries (see the following subchapters).

3.2.1 Sweden

In Sweden, the government has been publicising various school performance data online for almost 20 years. In 2001, the Swedish National Agency for Education [Skolverket] launched a web portal called *SIRIS* [Skolverkets Internetbaserade Resultat- och kvalitets Informations System], which contains various information on pupils' learning results and other school indicators (e.g. the municipalities' annual quality reports, leaving certificates from the ninth year of compulsory school, results from national tests and structural facts such as pupil-teacher ratios, teachers' qualifications, costs etc.) Another official statistical tool *SALSA* [Skolverkets Arbetsverktyg för Lokala SambandsAnalyser] takes different background variables (e.g. parents' educational level) into account in its analyses⁹.

The function and the target group of the *SIRIS* web portal was defined on its original website as follows [translation by the author]:

With SIRIS, we aim to make it easier for schools and municipalities to see what can be improved by examining their own performance and comparing themselves with others ... SIRIS is intended primarily as an aid to everyone who works in schools, to pupils and parents, and to municipal employees and politicians. The Agency wants SIRIS to be a tool for schools to use in their quest for improvement.

...

The key social function of schools means that citizens have a democratic right to have access to this information. Childcare and education affect almost everyone. In the Agency's view, public access must therefore be as extensive as possible. (Skolverket n.d.; see also Article I)

More recently, in 2015, the agency launched a new website (Skolverket n.d.) to help to investigate school evaluation data¹⁰. Here, the user may select various indicators under closer scrutiny and also compare schools with each other. The short invitation text on the website summarises its main function [translation by the author]:

⁹ In addition, the local level organization, 'The Swedish Association of Local Authorities and Regions' as well as the two teacher unions, 'Lärarförbundet' and 'Lärarnas Riksförbund', and the union for independent preschools and schools 'the Swedish Association of Independent Schools' have published reports on performance indicators and evaluation results (e.g. Friskolornas riksförbund 2017).

¹⁰ Note: The website was removed in 2019 after this chapter had been written. A new website with similar content can be found at .

Välja Skola [Choose the School] presents various indicators by the National Agency for Education that measure the schools' quality and results. Here you can search for and compare schools. The purpose of the website is to provide pupils and their parents with support when choosing a school. (Skolverket n.d.)

Välja Skola ÖPPEN SKOLINFORMATION

Sök Om skolverket Mina favoriter (0)

Sök på skolnamn

Sök Rensa sökning

1. Vilken typ av skola?

Grundskola Gymnasieskola

Från årskurs
Alla årskurser

Till årskurs
Alla årskurser

Välj huvudman
Alla skolor

2. Var ska skolan ligga?

Adress Län & Kommun

Sök skolor nära denna adress

Sök plats Dela plats

Hur nära?
0 km 100 km 200 km

Här kan du läsa om regler för val av kommunal och fristående skola

3. Vad är viktigt för dig?

☐ Avstånd

☐ Elever på skolenheten

☐ Elever per lärare

☐ Lärare m leg. o behörighet (%)

☐ Godkänt alla ämnen åk 9 (%)

☐ Meritvärde åk 9

4. Jämför skolor med varandra

Jämför Namn på skolenhet

<input type="checkbox"/>	Vasa International School of Stockholm/International section, Stockholm	♥
<input type="checkbox"/>	AGAPE-skolan, Hjälmare	♥
<input type="checkbox"/>	AL5/Falken, Västervik	♥
<input type="checkbox"/>	Abisko skola, Abisko	♥
<input type="checkbox"/>	Abrahamsbergsskolan, Bromma	♥

Figure 2. A screenshot of the Välja Skola [Choose the School] web portal in Sweden (Skolverket n.d.)

The function to promote information-based school choice is legitimated further on the website as follows [translation by the author]:

Choosing a school

All children and young people are entitled to a good school. Everyone also has the right to choose which school they want to go to, with certain restrictions that you can read more about under the <Rules for school selection> tab. The freedom of choice has made schools create profiles of themselves in several ways. They may want to invest in something they are particularly good at and of course also attract pupils.

The evaluations of the Swedish National Agency for Education and the supervisions of the School Inspectorate show that schools differ in terms of how well the national goals are met. The results of the national tests show differences

between the schools and in the pupils' grades. It may be due to a range of things. Often the pupils' background is highlighted as a factor that affects the results. But no matter which school the pupil goes to, it should have good teaching and be a safe environment for all the children. The school should help all the pupils reach as far as possible. It is a matter of equality and justice.

How to know which school is good? Getting an overall picture of a school requires that you find out the facts and put together different pieces of the puzzle. Above all, you should ask yourself what do you think that is important? The National Agency for Education's website 'Choose a School' will help you find various information about the teachers and the pupils' results on the schools you want to compare. There are also examples of questions that you can ask the school to learn more. (Skolverket n.d.)¹¹

What is important to note is that no government in Sweden, Denmark or in Norway has officially supported school ranking lists. On the contrary, the governments have continuously insisted that the indicators ought not to be used for school rankings [translation by the author]:

SALSA should not be used for school rankings. SALSA alone cannot answer the question of whether one school is better than another. How good a school's quality is, is an overly complex issue to be able to capture in a simple measure, whether it is a raw or model-calculated value. (Skolverket n.d.)¹²

However, this is an extremely complex issue to control. Once the performance indicators are publicised school-specifically, the results are easily sorted into a simplified ranking order in the media or by another third party.

3.2.2 Denmark

The basic education QAE framework in Denmark has gone through several phases of development work since the 1990s. In the 2000s, the role of national testing has been strengthened significantly. After many years of pilot testing and trial runs, national tests were implemented and launched officially in 2010. The national tests are mandatory, testing pupils from grades 2 to 8 in six teaching subjects. (Beuchert & Nandrup 2015.)

¹¹ Note: The website <http://valjaskola.se/artikel1.html> was removed in 2019 after this chapter had been written.

¹² [Original text in Swedish]: SALSA ska inte användas för att rangordna skolor. SALSA kan inte ensamt svara på frågan om en skola är bättre än en annan. Hur god kvalitet en skola har är en alltför sammansatt fråga för att kunna fångas i ett enkelt mått, vare sig det är ett faktiskt eller modellberäknat värde.

Various indicators of school performance are collected in an official web portal *Skoletal* under the Ministry of Children and Education [Børne- og Undervisningsministeriet]. Its main function is to provide information on the schools and their performance for the citizens [translation by the author]:

The purpose of the tool is to provide quick and flexible access to key statistics for primary schools, so that you can better compare and follow the developments in the schools. Here you will, among other things, be able to see how the students thrive at school and how they manage after the 9th grade exams. (Ministry of Children and Education n.d.)

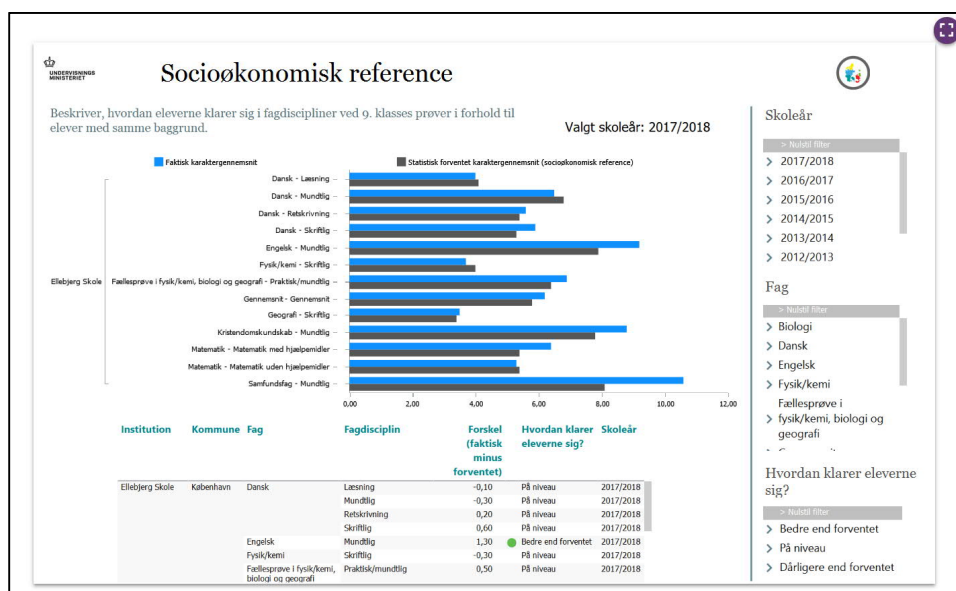


Figure 3. A screenshot of the Skoletal web portal in Denmark¹³

In Denmark, the policy guideline for the data publicising is somewhat complex. For example, the official guideline expresses explicitly that the results of national tests must not be used for any school rankings (see Ministry of Children and Education 2017, 12). Yet, the web portal allows the user to investigate the indicators school-specifically or to select up to three individual schools for a simultaneous comparison. These indicators entail for example grade point averages, number of pupils, pupil absence data, transition figures to secondary education, well-being indicators and national testing results. Despite the

¹³ The figure shows how the indicators of a randomly selected school are presented in comparison to schools with a similar socio-economic environment (e.g. på niveau [on the expected level]; bedre end forventet [better than expected]; dårligere end forventet [lower than expected]).

instructions, the data have been used for school rankings in the media (e.g. Søndagsavisen 2015).

3.2.3 Norway

In Norway, the comprehensive school QAE framework has been constructed and developed stepwise from the late 1980s. As in the other Nordic countries, the emergence of evaluation in education followed the changes in governance which emphasised decentralisation, management by objectives, output oriented policy and the use of performance measurements. However, only in the 2000s has Norway gradually established a national evaluation framework. It consists of various QAE practices, e.g. standardised national level pupil testing, school inspections, local level evaluations as well as schools' self-evaluations and well-being surveys (Elstad 2009; Norwegian Directorate for Education and Training 2011; Skedsmo 2011).

The need to strengthen evaluation measures was accelerated by the poor performance in the international assessments, most importantly in the PISA assessment. National pupil testing and the *Skoleporten* [the School Portal] web portal were developed as the first elements of the 'the National Quality Assessment System' [Nasjonalt kvalitetsvurderingssystem NKVS], introduced in 2004. The objective of the *Skoleporten* was to provide various types of school evaluation data for the various stakeholders, *'so that schools, school owners, parents, students and other stakeholders have access to relevant and reliable key figures for basic education'* (Norwegian Directorate for Education and Training 2011, 13).

The aim of the new national quality assessment system was to promote the system's accountability. By making the results from each school publicly available, it was expected to increase accountability at all levels, internally at the schools and through external pressure (Norwegian Directorate for Education and Training 2011, 16). However, the issue concerning the use of results and publicising policy involved a much more heated political debate than in Sweden. In spring 2005, the results of the national tests in 2004 gained a lot of attention in the press, leading to massive criticism from various directions. The new policy was criticised for many reasons, e.g. the fear of school ranking lists, increasing teachers' pressure and narrowing the teaching and learning objectives. In the general election in 2005, the Norwegian Labour Party returned to power after four years' opposition and the red-green alliance called a one-year time-out for the national tests, and that the results would not be publicised in *Skoleporten* at the school level but only at municipal and county levels. (Norwegian Directorate for Education and Training 2011; Hovdhaugen, Vibe & Seland 2017.)

After a few years elaboration work on the test validity and reliability and underlining simultaneously that the tests would function as only one element of

the National Quality Assessment System, the national tests were reintroduced in 2007 for reading, mathematics and English at grades 5 and 8 and 9. Still, the publicising policy confronted a confusing period, in which the results of individual schools were not available on Skoleporten, but based on the Norwegian Freedom of Information Act, the Directorate was required to provide the press and others access to the school level information on request. Consequently, the school results were publicised in the media on various websites on the same day that the Directorate publicised its results only at the municipal and county levels. The somewhat absurd situation finally led the Directorate to take a proactive role to define which results were also to be presented at the school level. (Norwegian Directorate for Education and Training 2011; Hovdhaugen, Vibe & Seland 2017.)

Oversikt	Fakta om opplæringa	Læringsmiljø	Læringsresultater	Ressurser	Gjennomføring
Her finner du tall og indikatorer fra utvalgte områder	Elevlar, lærarar, skolar	Elevundersøkelsen	Standpunkt karakterer	Undervisningspersonell	Overgangar
Realfagsbarometer		Elevundersøkelsen - deltakelse	Trinn 10	Økonomi	Ikke skolenivå
			Eksamens karakterer	Ikke skolenivå	
			Trinn 10		
			Grunnskolepoeng		
			Trinn 10		
			Nasjonale prøver 5. trinn		
			Nasjonale prøver ungdomstrinn		
			Nasjonale prøver - fritatt og ikke deltatt		

* Økonomi og overgangar finnes ikke på skolenivå.

Figure 4. A screenshot of the Skoleporten web portal in Norway (Skoleporten n.d.)

Constant elaboration work with numerous revisions has been the feature of the development of the Skoleporten website. The website now presents aggregated school performance data on five different indicator areas (Figure 4). The indicators are presented, with some exceptions, at the school level [translation by the author]:

School facts: Contains factual information about schools, such as the number of students and number of teachers.

Learning achievements: Overall achievement marks and examination marks in Norwegian, mathematics and English; Results of the national tests; Exemption and absence rates in the tests.

Resources: Teaching personnel, finance and materials. Teaching personnel indicators include e.g. the rate of full-time qualified teachers; Financial indicators consist of salary expenses and operating expenditures; Material indicators include

e.g. the number of computers per student and the number of computers connected to the internet.

Learning environments: Measured mainly through the Pupil Survey, which includes themes such as 'well-being', 'student democracy', 'physical learning environment', 'bullying', 'learning motivation' and 'academic guidance'.

Transition to / completion of upper secondary education (not at school level): Various transition and upper secondary indicators: e.g. percentage of pupils who continue directly into upper secondary education; percentage of pupils who complete upper secondary education in the normal length of time; percentage of pupils who did not complete upper secondary education, etc. (Skoleporten n.d.)

4 NEO-INSTITUTIONAL POLICY RESEARCH AS THE THEORETICAL FRAMEWORK

In the previous chapters I demonstrated how the features of ‘the global testing culture’ (Smith 2016), that is, increased accountability measures, standardised pupil testing and, above all, publicising of school-specific performance indicators, are present in the Nordic countries. Various statistical data and performance indicators have become more central in educational governance, and in Sweden, Denmark and Norway, official data are now collected, interpreted and transmitted to the public in the government’s internet web portals.

How is it to be explained that Finland has not followed the other Nordic countries’ publicising policy and practices, despite the many similarities in their political and administrative culture? This is the primary question that instigated this research. An easy answer to the question would perhaps be sufficient – it is just not ‘our thing’; it is not in our culture. Yet, in order to formulate a more analytical explanation, a closer look into the processes is needed, one in which the established practices and rules, the policy discourses and their underlying rationalities have evolved and how the legitimacy of these institutional elements are either supported or challenged.

In this chapter, I have presented the theoretical frame of my research, which applies and utilises the theories and the concepts developed under neo-institutional policy research. I will first discuss briefly the research streams of neo-institutionalism (4.1) and after that, I will present the writings of Vivien A. Schmidt (Schmidt 2008, 2010; also Carstensen & Schmidt 2016) on the newest research stream, *discursive institutionalism*. The theoretical framework of this research is illustrated in the last subchapter (4.3). In this research, the decline in the Finnish PISA results since 2009 is understood as a potential ‘critical juncture’ or turning point, in which new ideas and discourses may arise to challenge the legitimacy of the Finnish policy and its principles.

4.1 The different schools of neo-institutionalism

Neo-institutional research (also new institutionalism) began to draw more attention among policy researchers from the late 1970s (e.g. DiMaggio & Powell 1983; March & Olsen 1989; Meyer & Rowan 1977; Powell & DiMaggio 1991). Neo-institutionalism is often described as including three separate but often overlapping schools of thought or research branches: historical, rational choice and sociological institutionalism. More recently, discursive institutionalism has been described as the fourth and the newest research branch (see Chapter 4.2). Each research approach has its own concepts and approaches in its explanatory

models. However, common to all the branches is that they seek to elucidate the role that institutions play in the determination of social and political outcomes. Thus, the institutions are not simply seen as a product that results from a policy process under the given power relations. Instead, institutional analysis explains how institutions affect the behaviour of the actors – how do actors behave, what do institutions do, and why do institutions persist over time the way they do? (Hall & Taylor 1996; Hay 2006; Lecours 2005.)

In this research, I understand the policy relating to publicising the school performance indicators as being an institutionalised social practice. This includes historically, culturally and discursively established formal and informal rules and practices, in which who in society are entitled to access the information, how and why, is defined. As shown before, the opposite publicising policy solutions lead one to think that the benefits and the disadvantages of the school-specific publicising policy in Finland must be constructed differently from the other Nordic countries. Similarly, it can be presumed that the idea of citizens being entitled to or needing such information is not perceived in the same way in Finnish society as it is in the other Nordic countries.

The abovementioned first three neo-institutionalist research streams have been more prominent in explaining the permanence of institutional practices than their change. Looking from the perspective of ‘rational choice institutionalism’, it could be argued that in the 21st century, the Finnish PISA success has downplayed the reasons for a calculative and rational policy actor to suggest fundamental changes in the Finnish basic education QAE policy – in other words, ‘if it’s not broken, don’t fix it’ view.

‘Sociological institutionalism’ highlights the meaning of the values, norms and habits behind the institutional permanence. The institutions tend to create and follow the ‘logic of appropriateness’ (March & Olsen 1989), which works to resist alternative solutions. Here, it could be argued that a demand for school performance data to be publicised, whether expressed by a policy actor or a citizen, is recognised to be in conflict with the normative behaviour institutionalised in the Finnish QAE culture and wider in society. It could be argued that this view resonates well with research that has shown the high level of institutional trust in Finnish society. However, this is not very different from the other Nordic countries (e.g. Kouvo 2011; Listhaug & Ringdal 2007; Marien 2011).

‘Historical institutionalism’ highlights the meaning of past events. Previous decisions, actual events and their interpretations tend to set limits (or opportunities) for peoples’ behaviour in later events, often understood as path-dependent trajectories. Even though path-dependency must not be understood as deterministic, previous acts shape the possibilities between future choices. The historical institutionalism and path-dependence can be thought to have two variants that explain the permanence of the institutions (Mahoney 2000). The first

variant overlaps with the rational choice institutionalism, emphasising a sort of cost-benefit-rate between permanence and change: *‘With increasing returns, an institutional pattern – once adopted – delivers increasing benefits with its continued adoption, and thus over time it becomes more and more difficult to transform the pattern or select previously available options, even if these alternative options would have been more “efficient”’* (Mahoney 2000, 508). The second variant in historical institutionalism comes closer to the sociological institutionalism and sees the historical evolution and institutional trajectories as a chain of temporally ordered and causally connected events in which habits, values and practices are carried on. Thus, a change in the trajectory needs a certain, often contingent event or external factor that works as a critical juncture.

4.2 Discursive institutionalism – From ideas to discourses

While the older research streams of new institutionalism have been more prominent to explain the permanence of institutional practices, the aim with the fourth and the newest school, *discursive institutionalism*¹⁴, has been to shift the focus to the relevance of the ideas and discourse behind the institutional change. For Schmidt (2008; 2010; also, Carstensen & Schmidt 2016), ideas – understood here as interpretations or patterns in which the actors give meaning of the experienced world – make the substantive content of a policy discourse. Ideas exist at three main levels of generality: the ideas occur as 1) ‘policies’ (or policy solutions) providing suitable means for solving a specific problem or achieving the objectives set; 2) as ‘policy programs’ that underpin the policy ideas and reflect their underlying assumptions. These may be treated as ‘paradigms’ (Hall 1993), ‘programmatic beliefs’ (Berman 1998) or ‘problem representations’ (Bacchi 2009). Or, as Schmidt puts it:

These programmatic ideas are at a more basic level than the policy ideas because they define the problems to be solved by such policies; the issues to be considered; the goals to be achieved; the norms, methods, and instruments to be applied; and the ideals that frame the more immediate policy ideas proposed to solve any given problem. (Schmidt 2008, 306.)

At the third and even deeper level, the ideas occur as 3) ‘philosophies’, as worldviews that undergird the policies and programs with organising ideas, values and principles of knowledge and society. Whereas the policy and the program ideas make the surface for debate and policymaking, the philosophical level works as the fundamental core, often remaining implicit or taken for granted. (Schmidt 2008.)

¹⁴ Also called as Constructivist institutionalism (see Hay 2006).

For Schmidt (2008), the policies, policy programs and philosophies contain two types of ideas. Whereas the ‘cognitive ideas’ tend to construct ‘what is and what to do’, the ‘normative ideas’ attach values to political action by guiding ‘how things ought to do’ and aiming to legitimate the policies in a program through reference to their ‘appropriateness’ (see March & Olsen 1989) in the society.

Now, the ideas become conveyed, challenged and exchanged in the discursive interaction processes between different actors and audiences. The policy discourse, understood here as institutionalised structures of meaning that channel political thought and action in certain directions (see Connolly 1983), becomes shared and interpreted in different discursive forms – in policy justifications and arguments, official documents, public debate, narratives, collective memories and more. Here, Schmidt (2008) points out the need to separate the *coordinative discourse*, produced mainly by the central policy actors, and the *communicative discourse*, which refers more broadly to the different actors (individuals, the media, think tanks etc.) in society:

In the policy sphere, the coordinative discourse consists of the individuals and groups at the centre of policy construction who are involved in the creation, elaboration, and justification of policy and programmatic ideas. These are the policy actors – the civil servants, elected officials, experts, organised interests, and activists, among others – who seek to coordinate agreement among themselves on policy ideas. (Schmidt 2008, 310.)

The arrows of discursive interaction often appear to go from the top down. Policy elites generate ideas, which political elites then communicate to the public. Political elites often interweave the coordinative and communicative discourses into a master discourse that presents an at least seemingly coherent political program. The master discourse provides a vision of where the polity is, where it is going, and where it ought to go. (Schmidt 2008, 311.)

In a more recent article (Carstensen & Schmidt 2016), Schmidt points to three types of exerting ideational power: power *through*, *over* and *in* ideas. These three types of ideational power go along with the previous three levels of ideas (policy, policy program and philosophy) described above. Important for my research topic is the view that the political nature of the policy fades when reaching the deeper level of ideas:

That is, while the other forms of ideational power are focused more directly on the interaction going on between ideational agents, power in ideas concerns the deeper-level ideational and institutional structures that actors draw upon and relate their ideas to in order for them to gain recognition among elites and in the mass public. In this perspective, power in ideas concerns the ways that agents seek

to depoliticise ideas to the degree where they recede into the background, meaning that they become so accepted that their very existence may be forgotten, even as they may come to structure peoples' thoughts about the economy, polity and society. This may, for example, happen as policy programmes become taken-for-granted in terms of their methods, instruments and goals such that they, too, fade into the background. (Carstensen & Schmidt 2016, 329.)

4.3 Critical juncture, ideational change and institutionalised elements

Figure 5 illustrates the theoretical framework or model in this dissertation, utilising many of the abovementioned concepts of the overlapping research branches under neo-institutional policy research.

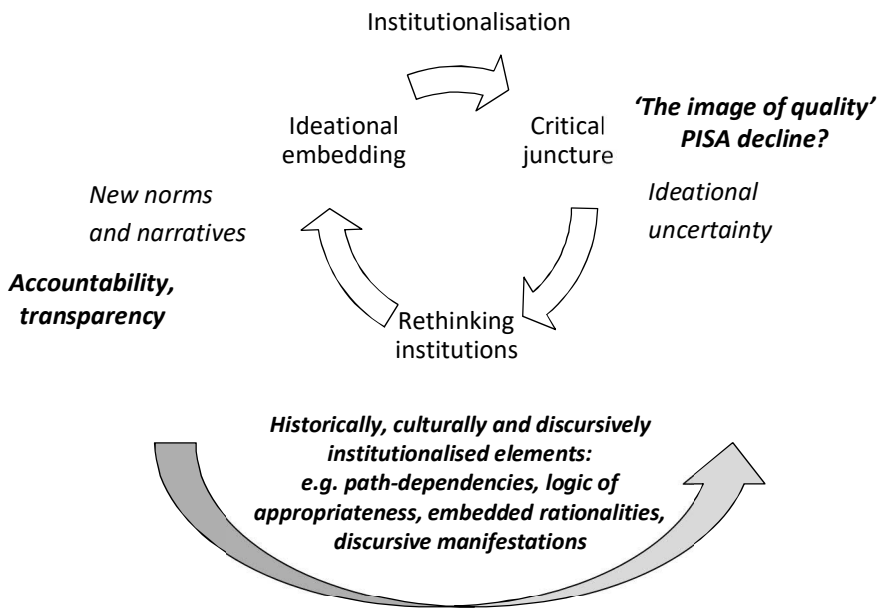


Figure 5. Theoretical framework in this dissertation (inspired by Erkkilä 2010)

The current publicising policy and its practices are understood to have evolved in an institutionalisation process in an environment that is constantly potentially open for change. Occasionally the policymaking environment faces a potential 'critical juncture', an event or factor that may open up space for change. A critical juncture does not have to be a revolutionary event but rather as Capoccia (2015, 147–8) puts it, *'a moment in which uncertainty as to the future of an institutional arrangement allows for political agency and choice to play a decisive causal role*

*in setting an institution on a certain path of development, a path that then persists over a long period of time.*¹⁵

The international large-scale assessments, foremost the OECD's PISA assessment, have become important references for policy actors to argue for policy reforms in the basic education systems in the first few decades of the 21st century (Meyer & Benavot 2013). In Sweden, Denmark and Norway, the implementation of the intensified evaluation measures and the web portals with various indicators preceded a heated debate on the quality of the comprehensive school systems and the PISA results. In Finland, the PISA success has supported the legitimacy of the basic education policy. However, since 2006 and more visibly since 2009, the Finnish PISA scores (Figure 6, top line) have declined continuously, approaching the level achieved by the other Nordic countries. Thus, following these notions, the decline in the Finnish pupils' PISA scores since 2009 is taken in this model, if not as an actual 'critical juncture' but more as a potential turning point, in which 'ideational uncertainty' on the prevailing institutional practices may start to occur. In that case, the current institutions become critically observed and 'rethought', and 'new norms and narratives' may start to arise and challenge the previous ones.

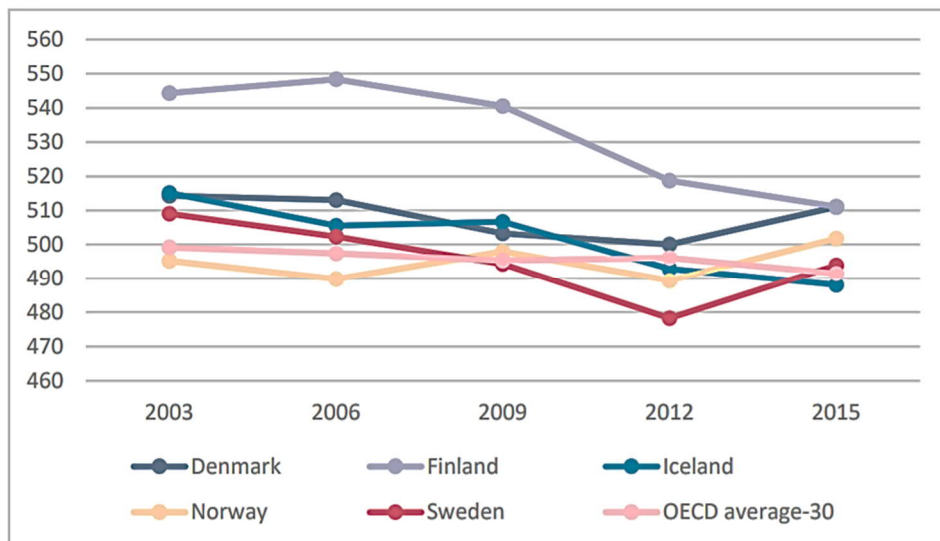


Figure 6. Mathematics trends in the OECD's PISA assessments in the Nordic countries (Nordic Council of Ministers 2018, 19.)

¹⁵ For other definitions, see Capoccia & Kelemen 2007; Mahoney 2000; Pierson 2000; Thelen 1999. As the critical junctures are often in their nature more or less contingent, the view on ideational change comes very close with the idea of policy actors' capability to 'play with the contingency', see e.g. Simola, Kauko, Varjo, Kalalahti & Sahlström 2017; Kauko, Rinne & Takala 2018.

In this research, ‘accountability’ and ‘transparency’ are understood as the core concepts, through which the change becomes discursively manifested. Eventually, if the political circumstances are advantageous, the new norms and narratives may become ‘ideationally embedded’ and result finally in new ‘institutionalised’ policies, practices, policy discourses and actor behaviour.

However, ‘the historically, culturally and discursively institutionalised elements’, meaning here mainly the previous policies, established practices and their embedded rationalities, may provide either advantageous, but more often resistant platforms for the ideational change. Thus, in this view, the arising ideas become conveyed, challenged, and exchanged in the discursive interaction processes between the present day actors and audiences, but with the past rationales also being challenged. Furthermore, it can be reasoned that if the institutionalised elements have become deeply embedded in society as taken for granted, it is more likely that the ideas that are arising continue to be resisted and buffered than generated further into changing policies and institutional practices.

To remind the reader, when starting with my dissertation project in 2013, there was not yet a clear understanding of how the QAE policy discourse in Finland was about to evolve with the Finnish PISA decline. Thus, whether it could be labelled as a critical juncture was yet to be explored but was eventually done so in Article III. In Chapter 5, I will move to describe the conduct of my research in more detail.

5 ON CONDUCTING THE RESEARCH, GATHERING DATA AND APPLYING THE METHODOLOGY

In Chapter 5, I have elucidated in more detail about the conduct of my research. This dissertation has been far from a straightforward research process, but more like a journey that has evolved gradually, along with contingent elements and changing paths. In the first subchapter (5.1), I have described how my research setting has originated and evolved during the research process. After that (5.2) I have outlined the research questions, which follow closely to those expressed in each of the three original sub-studies (Articles I, II and III). In 5.3, I have reported the empirical data and discussed my methodological choices in this research. The final subchapter (5.4) consists of reflection on the research ethics and validity.

5.1 Evolving of the research setting

I got interested in my research topic during my master's degree, in which I had the opportunity to familiarise myself with two interview data sets with Finnish educational experts, the EGSIE project in 1999 (Lindblad & Popkewitz 1999) and the FabQ project in 2006–9 (Ozga, Dahler-Larsen, Segerholm & Simola 2011). My observations on the Finnish QAE policy and especially on the publicity policy that was discussed in a clearly negative manner, guided my interest to scrutinise the topic further both theoretically and empirically.

I was motivated by questions like the following: Why is school ranking policy opposed in Finland so strongly? If the consequences of school ranking policy are as detrimental as is expressed in Finland, why is such a policy practised in many other countries? Above all, how can it be explained that our closest neighbouring reference countries, the other Nordic countries, have all ended up with an opposite policy solution in their publicising policy? How is it possible that publicising school-specific performance data in the Swedish society is a legitimated practice, while in Finland it is heavily opposed and condemned?

Article I (Chapter 6.1) was based on this origin. In Article I used the Model of Argumentation by Stephen E. Toulmin (1958) to analyse how the Finnish educational experts justify the publicising policy in contrast to the policy arguments in Sweden. The interview data were collected in the FabQ project in 2007–8, so it was not surprising that the PISA success in Finland framed much of the argumentation and strengthened the legitimacy of the current policy guidelines.

Yet, the fact that the decisions for the current publicising policy were developed in both Sweden and Finland during the 1990s, that is before the first

PISA assessment in 2000¹⁶, made me realise the importance of historical elements. I was convinced that the current, opposing policies must somehow resonate with the earlier structures and policy rationalities in both countries. This is how I came to scrutinise and contrast the historical institutionalisation of the QAE policies in Finland and Sweden. In Article II (Chapter 6.2), I took a wider historical perspective in order to understand the contextual roots and events in which the QAE practices and especially the national testing framework had been developed in Sweden and Finland.

More or less by chance, the research context of my dissertation was extended to cover Denmark and Norway¹⁷. I was involved with a comparative research project (DYNO) in 2017 in which I had a great opportunity to familiarise myself with interview data collected in Norway and Denmark. By this extension, the Finnish ‘upstream policy’ in my research began to show even more anomalies in the Nordic context. The case was no longer contrasting Finland only with Sweden, in which neoliberal policies (e.g. the marketisation of education, intensified accountability and increased QAE methods) have been perhaps the most visible within the Nordic region (e.g. Lundahl, Erixon Arreman, Holm & Lundström 2013), but also with Denmark and Norway, who have followed Sweden in their publicising policy with official web portals.

With the new empirical interview data, my research focus shifted towards thinking about the potential for policy change. In Article III (Chapter 6.3) co-authored with colleagues Sara Juvonen, Petteri Hansen and Janne Varjo, I approached the interviews from the theoretical framework of discursive institutionalism. In Article III, I pushed the concepts of ‘accountability’ and ‘transparency’ closer to centre stage in my research.

In my view, this coincidental opportunity to join the DYNO team and access the newly collected empirical interview data from all the Nordic countries, only deepened my original research setting on contrasting Finland with Sweden. Each sub-study progressed my research by providing one new consequential research task after another. In this Summary, these individual sub-studies (Table 4) are brought together to form a coherent whole.

This dissertation is positioned within the research area of comparative policy research in education. In my research, I understand the publicising policy of school performance indicators as a social practice that has been institutionalised in each society within a unique socio-historical process embedded with historically, culturally and discursively established rules, practices and rationalities. In this sense, this research could be described as multidisciplinary, applying from the research fields of policy research, sociology and history,

¹⁶ The first PISA results were published in November 2001.

¹⁷ In the DYNO research project, interviews with policy actors were also collected in Iceland, but for practical reasons concerning my research schedules, I decided to leave Iceland out of my research.

touching upon educational policy, governance, policy discourses and the processes of policy legitimization.

Table 4. Research setting in the original sub-studies (Article I, II and III)

	Focus of analysis on	Theoretical framework	Methodology / Contrasting	Data
Article I	Policy arguments, justifications	WPR model	Argumentation analysis (Toulmin model) / Finland – Sweden	7 Finnish FabQ interviews, 2007-08
Article II	Established policies and practices, path dependencies	Historical institutionalism	Historical analysis / Finland – Sweden	Analytical literature review
Article III	Policy discourse, Narratives	Discursive institutionalism	Policy discourse analysis / Finland - Sweden - Denmark - Norway	58 interviews: Fin (9), Swe (17), Den (14), Nor (18), 2015-17

5.2 Research questions

The overall aim of this dissertation is to clarify how and why Finland has been able to resist the pressures of the ‘global testing culture’ and the idea of publicising school-specific performance results. Each original sub-study approaches the same research aim, yet with a specific research task. The research questions addressed in this dissertation are as follows:

1. How are the opposite publicising policies (being) justified in Finland and Sweden? (Article I)
2. How are the current publicising policies in Finland and Sweden explained through historically institutionalised path-dependent elements? (Article II)
3. How are the two core concepts that typically promote a school-specific publicising policy, transparency and accountability manifested in the policy discourses in Finland, Sweden, Denmark and Norway? (Article III)¹⁸

¹⁸ The fourth conceivable ‘research question’, ‘Does the decline in the Finnish PISA results since 2009 mark a critical juncture, in which new ideas or demands arise to challenge the legitimacy of the prevailing policy and practices?’, is understood more as a methodological question (see sub-chapter 5.3)

5.3 On data and methodology

This dissertation is positioned within the research area of comparative policy research in education. As stated earlier, the other Nordic countries were used in this research as a ‘reflective surface’ in order to make visible the particularities of the Finnish policies, policy discourses and its underlying rationalities. Thus, the comparative research setting that was present in the three sub-studies was an intentional methodological choice from the beginning. For this Summary, an overview of the official web portals was added (Chapter 3) in order to deepen the view and share knowledge on the different publicising policies and practices in the Nordic countries.

Methodologically, comparative policy research can be utilised in a range of ways (e.g. comparing countries, regions, institutions, processes etc.) It is essential for the researcher to be aware of his/her methodological choices. For example, the selection and the number of comparable countries may easily lead to different results and interpretations. Another relevant issue to note is how the similarities and the differences between the observable units should be interpreted. For Todd Landman (2008, 298), it is essential for the research problem, case selection, inferential aspiration and the theorising to make up a coherent whole.

In this research, the decision to compare and contrast the Finnish policy with policies in the other Nordic countries was natural: there are apparent similarities across Nordic societies, a strong emphasis on social and educational equality and the principle of publicity in their administrative culture provided a fascinating origin for comparative research. Having a ‘few-country’ study also enabled scrutiny of the publicising policy as a historically institutionalised process, an issue that is often lacking in ‘many-country’ studies (Landman 2008).

The main empirical research data I used were comprised of interviews with key policy actors in basic education policy. In this key informant technique (see Tremblay 1982), informants are considered to have specialised knowledge on the observable phenomenon by their experience, status or occupation. In both the FabQ and DYNO data, the interviewees represented the central organisations in basic education policymaking and governance, including national and local level decision makers (including former and/or in office Ministers of Education), key officials in the educational ministries and in other executive educational offices, as well as representatives from other relevant stakeholders such as the teacher unions, (economic) lobby groups and academic experts from the universities.

In the FabQ research project, the informants were selected by mapping the most relevant organisations in basic education QAE policy and then interviewing their central representatives. Seven thematic interviews with Finnish educational experts on QAE policymaking collected in 2007-08 were used in the analysis. Also, in the DYNO project, the key informants represented central organisations in basic education politics and governance. Here, the informants were identified

additionally through a ‘snowball sample’ method (also ‘chain sample’), in which the informants are asked to suggest other relevant informants (see Noy 2008). Fifty-eight interviews with educational experts in Finland, Sweden, Denmark and Norway collected between 2015 and 2017 were analysed. The interviews in both research projects took from 60 to 90 minutes and since all the interviews had been taped and transcribed, their re-use for my research was possible and convenient.

In Article I, I focused on analysing the justifications of the publicising policy. Using Stephen E. Toulmin’s Model of Argumentation (Toulmin 1958), I contrasted the policy justifications of the informants in the FabQ data with the official publicising policy arguments in Sweden to capture the structure of the argumentation logic. In Article II, the focus was on the historical institutionalisation of the publicising policy. Here, the literature review consisted of official documents and contemporary writings by senior experts and researchers in the field of basic education QAE policy in Finland and Sweden. In Article III, the analysis of the DYNØ interviews used policy discourse analysis in order to clarify the embedded rationalities behind the different publicising policy solutions within the Nordic countries.

In this research, the choice to focus mainly on the discursive manifestations of the educational experts followed the theoretical framework on discursive institutionalism and the division in the coordinative and the communicative policy discourse (Schmidt 2008). For Schmidt, the key policy actors aim to control the coordinative discourse in policymaking in their favour both by defining problems and by providing suitable solutions for the problems. Thus, the power in policymaking is not seen only as a struggle between different policy means, but moreover a struggle over who gets to name and define a specific issue as being ‘problematic’ (Bacchi 2009).

Thus, the basic idea to include an extensive set of interviewees from a range of organisations in the analysis was to capture the space of the coordinative policy discourse that may have consisted of various and conflicting voices and views. In addition, the choice to focus on the interviews instead of conducting a comparative analysis of official policy documents followed the theoretical framework and the idea of the potential critical juncture in this research. Before eventually ending in any formal policy document in Finland, it might be expected that the ‘new ideas’, the shortcomings of the current policy or needing to reform the QAE practices would somehow first become expressed and acknowledged in more informal discussions with the policy actors interviewed.

The interview analysis with the political elite is a special type of research method and communication practice (see e.g. Mykkänen 2001). The elite interviewees may be valuable informants for revealing inside information on the policy processes, which otherwise would not become visible. The level of specificity on how the questions are posed in the interviews affects the communication of the interviewees. Within this research, the interest was not so

much in the details. For example, the interview method used in the DYNO research project (Article III), in which the interviewees were posed only a single question to start the interview, *‘what is the story of your country’s basic education system?’*, suited my research interests. As it was in my interests to analyse how (or if at all) the publicising policy and its practices were problematised in the interviews, a broad interview set-up with no fixed interview themes was extremely useful.

As presented earlier, analysing two sets of interview data collected in different time periods, entailed a special aim to capture the ‘effect’ of the decline in PISA scores in Finland in the policy actors’ discourses. Even though the two data sets were not analysed with each other, the impression on both sets of data showed no clear change in the discursive manifestations. During the many years this dissertation project has been spread, I have naturally followed the public debate on the Finnish basic education system and QAE policy. Despite focusing on the coordinative policy discourse in this research, no real openings for policy change were witnessed in the wider communicative policy discourse (Schmidt 2008).

To inform the reader, in the early phase of my dissertation project, I personally conducted eight interviews on my research topic, mostly with Finnish educational experts, but I also visited Skolverket, the national evaluative agency in Sweden, where I interviewed two officials. Unfortunately, these interviews turned out to have only a preliminary role in my research, as I never managed to arrange times to collect data set that was sufficiently detailed to be used in a published research article. Yet, even if these data were never analysed systematically, my impression was not contrary to the FabQ and DYNO datasets. For example, to begin the interviews, many of the Finnish interviewees asked for my own motives and intentions with this research topic – which undeniably left me with the feeling of dealing with a delicate research topic. However, the impression in the interviews with the Swedish informants was totally the opposite, as one interviewee finally put it after my deep pondering about the Swedish publicising policy ‘well, here in Sweden, we are just used to publishing everything’.

5.4 On research ethics and validity

The opportunity to be part of an established research group (KUPOLI) has supported the conduct of this research in its different stages. First, the chance to analyse the two interview data sets from the FabQ and DYNO projects, has provided valuable benefits. In a broader sense, the organisational structures and practices of the research unit have provided a framework for conducting the research according to the ethical and responsible guidelines of scientific research (see The Finnish Advisory Board on Research Integrity 2012).

In qualitative research, especially when dealing with interviews, various research ethical issues must be considered. Here, I will point to three aspects,

which in my view are essential to reflect on: 1) the issue concerning the informants' anonymity, 2) analysis of interview material collected by other researchers and 3) the reflection relating to the nature of the research topic and the research findings.

Perhaps the most general problem in qualitative research with interview data touches the anonymity of the interviewees – to what extent can the anonymity of the interviewees be guaranteed without losing the contextual meaning or the informant's position in the interview citations. The practices on anonymising vary by research context and no clear, comprehensive instructions can be given. Yet, it is essential to have clearly expressed ground rules, in advance, on the aims of the research, the use of the interviews, the anonymity policy as well as the potential re-use of the data, to support the confidentiality between the interviewer and the informant (e.g. Ruusuvuori & Tiittula 2005).

Within this research, the guidelines for the informants' anonymity had already been defined in the research plans of the two research projects. For example, in the DYNO research plan, it was defined that *'All the research participants (policymakers and in the extended option also parents, teachers and pupils) will be kept fully informed about the research, its purpose and voluntary nature. ... All the data concerning parents, pupils, schools and their personnel will be confidential, and identifying characteristics will be modified or withheld in order to preserve confidentiality.'* In order to follow these guidelines, at the beginning of each interview, the interviewees were asked to sign a letter of consent, in which the aim of the research project, the matters concerning the identity of the interviewee, the voluntary nature of being interviewed and the storage of the interview material were agreed on.

In the sub-studies in this dissertation, all the interview citations were anonymised without mentioning the institutional status of the interviewees. Nor were names and occupations/institutions of the interviewees shown in the appendix. The coding used in Article III (e.g. FIN03, NOR09) expresses only the country and a running number of the interviewee. Such a decision is understood to weaken some of the explanatory power of the research findings but it was made to ensure that no violation of earlier agreements on anonymity were made by mistake, especially as the number of the policy actors interviewed is quite limited and some of the informants might be predicted by their organisation.

Another relevant methodological issue related to research ethics concerns the problem of analysing material that has been collected by other researchers. As mentioned earlier, I participated personally in only the last few interviews within the DYNO project. Thus, the analyses of both interview data sets were based on transcriptions without experiencing the non-verbal communication in the interaction process. However, the opportunity to discuss the research material with the actual interviewers – my supervisor and other colleagues, who were also

co-writers of the third sub-study – in my view supported the validity of my interpretations of the data.

However, within this whole dissertation, the most uncomfortable and contradictory ethical dilemma has concerned the research topic itself. By this I mean the sensitivity (or fragility, see Simola et al. 2009) that is entwined around the question on the publicising policy of the school performance in Finland. Indisputably, the legitimacy of the Finnish publicising policy nourishes from its depoliticised and taken for granted nature. The awareness that my research approach, to analyse critically the institutionalisation of this policy and its rationales and at the same time to contrast it with opposite policy practices in other Nordic countries, may entail elements that politicise the topic or raise it into a wider public discussion, has been a personal contradiction. In the worst-case scenario, drawing attention to other policy options in the Nordic countries might open a Pandora's box by causing a series of unfavourable changes in the Finnish policy discourse and practices. Yet, this kind of ethical thinking relating to the research's consequences in society is always present, especially in the nature of social science research. Even if the researcher cannot be held responsible for the instrumental or political use of his/her research or its findings, at least the researcher should be aware of the political and ideological role of social science research (see Alvesson & Sköldberg 2009; Kauko, Gorodski Centeno, Piattoeva, Hinke Dobrochinski Candido, Gurova, Medvedeva, Santos, Suominen & Zhou 2018).

6 THE RESULTS – EXPLAINING THE FINNISH PUBLICISING POLICY THROUGH HISTORICALLY AND DISCURSIVELY INSTITUTIONALISED ELEMENTS

In this Chapter, I will present a summary of each of the three original sub-studies included in this PhD dissertation. All three research articles went through a scholarly peer-review process and were published in an academic journal or edited book. As presented in the previous chapter, the research setting of these articles has been developed and shaped during the research process, to answer consequential research questions that were raised on the findings of each previous sub-study.

Together, the individual articles form a coherent whole, in which the institutionalisation of the Finnish quality assurance and evaluation (QAE) policy, culture and practice is scrutinised with a special focus on the publicising policy of school performance results. Methodologically, all the three articles represent a comparative policy research in education, in which the Finnish case is contrasted with the other Nordic countries.

In Article I, I scrutinised how the opposite publicising policies are justified in Finland and Sweden. In Article II, I approached my research topic from a wider historical perspective to examine how the current publicising policies in Finland and Sweden can be explained through historically institutionalised path-dependent elements. In Article III, I aimed to answer a question about how the two core concepts that typically promote school-specific publicising policy, transparency and accountability, are manifested in the policy discourses in Finland, Sweden, Denmark and Norway? The interview data analysed for this final sub-study is temporally positioned at a time when the Finnish PISA results have started to decline, understood in this research as a potential ‘critical juncture’ for discursive change in the Finnish QAE policy.

6.1 Tracing the argumentation logic of the opposite publicising policies of school performance results in Finland and Sweden (Article I)

In Article I (*Wallenius, T.J. (2015). Justifying opposite publication policies of school performance results in Finland and Sweden*), I focused on scrutinising how the opposite publicising policies are justified in Finland and Sweden. The empirical data consisted of seven thematic interviews with Finnish educational key experts collected in 2007–08 within the FabQ research project. The Finnish policy justifications were contrasted with the policy justifications in Sweden, e.g.

with official texts of the responsible evaluation organisation Skolverket and the Swedish FabQ interviews (see Segerholm 2009). A comparative research setting was used to make the construction of the argumentation logic more visible. Two research questions were addressed: first, how are the opposite publishing policies on school-specific performance results justified in Finland and Sweden?; and second, What ‘problem’ were the policies aimed to answer? Applying Carol Bacchi’s ‘What’s the Problem Represented to be?’ WPR theory (Bacchi 2009), in which policymaking is seen as a social process of ‘problematization’, the aim of the research was to analyse the policy justifications and to discern the policymaking context in which the ‘problems’ are constituted and the policy solutions and their justifications become rationalised.

Methodologically, the analysis applied a scholarly classic: Stephen E. Toulmin’s Model of Argumentation (Toulmin 1958), in which all three essential factors of policy argumentation were taken into account: the data (or premises; for Bacchi the ‘problem’), the warrants (the justifications, the arguments) and the claim (the policy solution) (Figure 7).

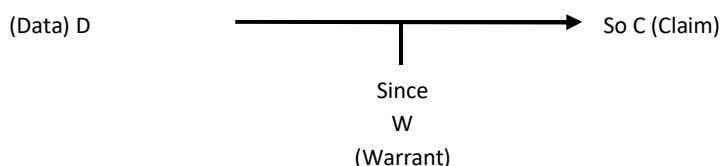


Figure 7. Stephen E. Toulmin’s Model of Argumentation (Toulmin 1958, 92).

The Toulmin model provided a suitable methodological structure under which to conduct the argumentation analysis stepwise. As for the first research question on policy justifications (in Toulmin’s model, the Warrants, W) the two examples below are illustrative to show the differences between the countries. First, an interview citation of a Finnish informant captured aptly the deep aversion to school ranking policy, embedded in the Finnish school evaluation culture [translation by the author]:

Now the ranking-list was the word we opposed ‘til the end, it actually was the specific justification in the decision for conducting sample-based evaluations...Like this one case, [refers to a piece of news on television], showing that this and that one is a poor school, as if it was only information for the parents... (FabQ interview)

This rejective view referring especially to the role of media was widely shared across the data yet expressed in several ways. In addition, the interviewees pointed out economic arguments, meaning here the cost-efficiency of the sample-based testing in comparison to a testing system to cover all the pupils and schools. Also,

the concern over teachers' autonomy or test-oriented teaching was expressed. However, the main justification in the Finnish argumentation model was the aim to control the acceleration of social segregation. This unwanted and detrimental trajectory was perceived as one that would intensify automatically if the results were publicised at the school level and the neighbourhoods with lower school performance results come to be avoided through a 'naming and shaming' process.

Now, when contrasting the Finnish justifications with the Swedish policy guidelines and policy justifications, a clear difference was noted. For example, on the Skolverket web portal SIRIS, the relationship between the government as an information provider and the citizens as the information users followed a significantly different logic and accountability hierarchy than in Finland:

The key social function of schools means that citizens have a democratic right to have access to this information. Childcare and education affect almost everyone. In the Agency's view, public access must therefore be as extensive as possible. (Skolverketa n.d.; see also Article I)

In Sweden, the decision to publicise performance results school-specifically through the government's web portal from 2001 entailed at least two intertwined justifications. First, it followed in line with the fundamental principle in Swedish governance, the idea of supporting citizens' democratic rights through transparency and public information, in which there is no exception for the school evaluation data. Secondly, the decision suited well the other market logic policy reforms that featured in the Swedish basic education politics in the 1990s – the introduction of a quasi-market school system, increased competition, parental choice and an intensified system of accountability through pupil testing and school publicity. Within this quasi-market context, the policy solution to provide all families with an equal opportunity to access relevant school performance data at the school level, becomes justified and logically coherent.

Finally, to answer the second research question, the adjacent Toulmin models in both countries were completed when looking at the societal context, in other words the premise (Data, D) or 'problems', which the opposite policy solutions aim to solve or answer. Here, I pointed out the image of the quality of the school system as the main contextual premise for both policy solutions. The heavy critique that had been directed at the Swedish school system since the 1980/1990s multiplied and continued in the 2000s along with the Swedish 'PISA shock'. Thus, 'the problem', the image of low quality and differentiated school performance, was to be fixed by the abovementioned policy reforms – among them the new publicising policy that allowed the school comparisons. In contrast to Sweden, the context of the basic education policymaking in Finland was not framed in the 1990s by a similar image of low quality. Eventually, the PISA success in the early

2000s strengthened the legitimacy of the Finnish comprehensive school policy, including the QAE policy and practices.

The first sub-study showed how the countries differ not only by their publicising policy, but also by their justifications and the contextual frame, the ‘problematisation’ (Bacchi 2009). Most importantly, the analysis showed a clear difference in the argumentation logic relating to the publicising policy. Whereas the Finnish rationality followed the idea of ‘quality through preventing inequality’, the Swedish rationality aimed to ‘raise the quality through competition’, typical for the neo-liberal ideology. Methodologically, this sub-study showed the feasibility of the Toulmin model in comparative policy research – something that at least I had not witnessed before.

Yet, the analysis and the findings in this sub-study raised new research tasks. First, the unison among the Finnish policy actors was baffling. The resistance towards the school rankings (noted also in previous research on the Finnish QAE culture, see e.g. Simola et al 2009), was widely shared in my data. Obviously, as the interview data were collected during this PISA success period, no major demands for policy reform were even expected in the data. Still, the absence of any comments concerning the opposite justifications in Sweden – equal access to information, transparency or democracy – was extremely interesting. Which elements of Finnish society has made it possible for such a strong consensual and self-evident view on the publicising policy to evolve in the Finnish basic education QAE policy and governance? Secondly, how does this view keep on sustaining its legitimacy? What will happen, if the ‘image of the quality’ of the Finnish school system weakens? The following two sub-studies continued with these questions by scrutinising the institutionalised elements of my research topic.

6.2 Explaining the current national testing practices in Finland and Sweden through historical institutionalism and path-dependencies (Article II)

In Article II (*Wallenius T. (2016). National Testing of Pupils in Finland and Sweden in Light of Historical Institutionalisation [originally published in Finnish]*), I approached my research topic from a wider historical perspective and scrutinised the historical development of the national level pupil testing policy and practices in Finland and Sweden. As presented earlier in this Summary, during the 1990s the countries had ended up practising significantly different QAE policies. Yet, to understand these different policy guidelines thoroughly, it became clear that I needed to observe their historical development.

In Article II, the research task was to explore how the current publicising policies in Finland and Sweden can be explained through historically institutionalised path-dependent elements. Theoretically, Article II represented

most clearly the research tradition of historical institutionalism (e.g. Pierson 2000; Mahoney 1999). Thus, my aim in this sub-study was to explore how the previous decisions and historically established practices and discourses have set certain path-dependencies to shape the later QAE policy trajectories in both countries.

The literature review started from the origin of the national level pupil testing in both countries. The comparison showed that the pupil testing in Sweden has had significantly longer historically-institutionalised roots than in Finland. Overall, Sweden has been one of the global forerunners in the development and the use of standardised testing in the schools. Even in the early 1940s in the time of the parallel school system and 20 years before the comprehensive school reform in 1962, the Swedish government introduced a standardised test, 'Standardprov', which in my view has an essential role in the institutionalisation process. The function of the test was to help the teachers to assess the pupils' grading, by which the pupils were eligible to apply for further studies. Most importantly, it was argued that the test supported equality (in the sense of fair grading) among the pupils (and families) and despite being voluntary, the test was broadly used and accepted by the teachers across Swedish schools for many decades (Ljung 2000; Lundahl 2009).

In Finland, the national testing started to have institutionalised forms notably later, during the 1960s/70s within the comprehensive school reform. From the very beginning, it was strictly defined what purposes the pupils' learning tests [in Finnish *koulusaavutuskokeet* or *yhteiset kokeet*] were designed for – not to have a nature of control over teachers or schools nor any role as a final exam (Saari 1983). Thus, the national testing framework never reached the level of institutional comprehensiveness within Finnish schools or educational governance as it did in Sweden, but was featured by low budgeting, unsystematic strategy and thematic, sample-based testing until the 1990s.

What is noteworthy is that the publicising of the test results did not have a significant role before the 1990s in neither countries policymaking. Yet, as the role of evaluation was strengthened during the 1990s in educational governance, the question of the use of the test results also came to be decided on. Here, the historically-established practices and rationalities on national level pupil testing were path-dependent elements on which to construct the platform for the decision-making in the 1990s.

In Sweden, the national testing was ordered obligatory in 1997/98 in order to evaluate systematically the quality of all the schools in the decentralised school system and to continue to ensure the fairness of teacher assessment. The national testing now covered all schools and the results were collected and publicised on the official web portals only few years later in 2001.

The publicising question did not become a key policy issue in Sweden. Within the QAE policymaking, the debate touched more upon the amount of pupil testing or school inspections. The publicising of school-specific performance followed

the principle of publicity and data publicising was legitimated by enhancing the citizens' equality, democratic right to access official data and by supporting the parental choice.

In Finland, on the contrary, the publicising issue took a much more decisive role as the Finnish QAE guidelines were formulated during the 1990s. It was decided to conduct national level pupil testing according to a sample-based method, above all for the reason that the method did not force the government to publicise the test results at the school level. Thus, the sample-based testing system eventually turned into a methodological gatekeeper to ensure and define for whom the evaluation results were to be addressed.

To sum up, based on the comparative analysis I pointed out three or four intertwined characteristics in the evolution of the national testing systems prior to the 1990s, which in my view shaped the later QAE decisions in Finland and Sweden as path-dependent elements. The long historical experience (*length*), the wide use of the tests among the teachers and the pupils (*breadth*) as well as the comprehensive nature of using the standardised national testing (*systemacy*), together explain the long, established and institutionalised role that the national level testing framework has had in the Swedish school evaluation culture. In Finland, national-level school evaluation did not reach a similar central position in the comprehensive school policy, which obviously enabled the Finnish policy actors to define the guidelines and to formulate the sample-based national testing and publicising policy differently during the 1990s.

Article II was concluded with a discussion that reflected some contemporary policy debates. For example, in Finland, the equivalence of teacher assessment was once again questioned, now especially with the findings presented in Najat Ouakrim-Soivio's dissertation (Ouakrim-Soivio 2013). In Sweden, the comprehensive school policy was one of the main themes in the 2014 parliamentary elections. The intensified QAE activities ordered by the Conservative governments in the 2000s were heavily criticised by the Social Democratic party. However, looking at it from a long historical perspective, the embedded rationalities concerning either national level pupil testing or the publicising policy, seemed to set boundaries and limit major policy reforms in both countries. In other words, both countries can be seen to have directed to different path-dependent trajectories in the early stages, a trajectory in which change becomes unlikely because of the institutionalised practices, rules and rationalities.

Yet, as described earlier, despite the different institutionalised restraints, the environment of the policymaking is constantly changing. New ideas and rationalities may appear in the discursive manifestations, which may start to challenge or even supersede the older ones if the circumstances are favourable for change. Thus, to investigate the elements of political change more closely, I moved my focus in the final Article III to analyse more recent policy discourses.

6.3 Understanding accountability and transparency in the Nordic QAE policy discourses (Article III)

In Article III (Wallenius, T., Juvonen, S., Hansen, P. & Varjo, J. (2018). *Schools, accountability and transparency - Approaching the Nordic school evaluation practices through discursive institutionalism*), I moved my focus from the historical analysis back to the more recent policy discourses. With my colleagues in the DYNO research project in the KUPOLI research unit, Sara Juvonen, Petteri Hansen and Janne Varjo, in the article we analysed the discourses of the educational experts on comprehensive school QAE and publicising policy, in four Nordic countries, Sweden, Denmark, Norway and Finland. The analysis comprised 58 interviews of central educational actors, collected for the DYNO project during 2015-17.

In the second sub-study, I had already pointed out how the Finnish publicising policy can be seen to entail historically institutionalised path-dependent elements, deriving from the time before the emergence of the global testing culture and public school performance indicators. The first sub-study showed how the Finnish publicising policy was justified and legitimated at the beginning of the 2000s with the more or less unexpected PISA success. Now, the aim of the final sub-study was to scrutinise whether the changing environment – referring mainly to the decline in Finnish PISA scores – would open up the policy discourse on the Finnish QAE policy and its guidelines. In addition, the new interview data enabled the Finnish QAE policy and governance to be contrasted not only with Sweden, but in a broader view within the Nordic region.

Theoretically, Article III utilised the writings of Vivien A. Schmidt (2008; 2010) on discursive institutionalism. The interviewees were seen as the key policy actors, whose opinions and views construct and shape the coordinative policy discourse space:

Political elites tend to interweave the coordinative and the communicative discourses into a master discourse which sets out not only the visions for policymaking – what is, and what ought to be – but also defines the terms and the frames for the public discussion – what is rationalist, appropriate, how and why (Schmidt, 2008). (Article III, p. 136.)

The interviews were built on a single main question, ‘*what is the story of your country’s basic education?*’. The open interview structure suited the theoretical framework in finding out which topics were raised and how. In our analysis, we focused especially on investigating the manifestation of the two interwoven concepts, accountability and transparency, through which the idea of collecting and publicising comparable school performance data are typically promoted in the educational policy and governance. Two research questions were formulated as

follows: 1) how is the recent national testing policy and especially the use of the results legitimated/challenged by the educational experts in the Nordic region? 2) which underlying ideas or programmatic beliefs seem to guide these discursive practices?

In general, evaluation was a central topic discussed throughout the data, especially when reaching more recent events in the interviews. However, the analysis showed significant country-specific differences between the interviews.

The Swedish interviewees discussed their national QAE policy in very practical terms – the argument supporting the need to have a systematic national testing system was to monitor the school system and to improve its overall quality. Even though the intensified accountability measures (such as increased pupil national testing) were causing worries for many informants, the publicising of the school-specific results was not challenged in the interviews. In our view, this indicated its institutionalised and naturalised nature. Transparency was seen as a tool for accountability, directed not only to the state but also towards parents, to support equal opportunities for parental choice.

Many Danish interviewees described the 2000s as a new era in the Danish school evaluation tradition, in which accountability measures were intensified intentionally. In the Danish interviews, the OECD had a central role as a driver of the new QAE policy and practices. The publicising policy was understood as being complex and conflicting, yet not unacceptable but rather as a feature of a modern information society:

So, I'm very sceptical about rankings. But I'm also extremely sceptical about the view that rankings should be prohibited. Because in a knowledge society, people who have a little bit of skills can publish whatever they like. It is no longer the idea that we have this one government that controls all the information. So, I would rather have that we undermine these rankings because we made them ridiculous, and we discuss them, and we compare them over the years. I would rather go in that direction, than having this idea that said 'you know, the people don't deserve to know because they don't have the competence to interpret these things' [...] I would say, publish it if you want to. Let's have the debate. And it's fairly ridiculous in the long run in a way. (DYNO interview, Denmark. Article III, p. 138.)

In the Norwegian interviews, the evaluation reforms and the implementation of the national QAE framework originated from the growing need to get systematic, research-based information on the school system. In the interviews, the idea of increased accountability in the comprehensive school system was acknowledged in terms of cost-efficiency or quality improvement, yet many times argued also that they would help and benefit the pupils themselves. As for the publicising policy, the difficulties in the implementation of the government's web portal (see

Chapter 3.2.3) were present in our data. For example, many interviewees were not able to describe exactly how the results were actually publicised. Yet, in general, the Norwegian approach resembled the Danish view – not prohibiting the data but informing that performance indicators are only one piece of information.

The Finnish interviews showed a clear difference with the other Nordic countries. Even though the decline in the Finnish PISA scores was noted as an alarming signal, no policy solutions that could be interpreted as measures to increase accountability was expressed in the data. On the contrary, the Finnish informants held a strong and a shared collective understanding of the aims and the mechanisms related to school evaluation. The low accountability measures were described basically as a case of common and mutual trust between all the different actors, including the teachers, school leaders, officials as well as the parents. Despite the growing concern on the fairness and equivalence in teacher assessment, introducing a more systematic evaluation system with standardised tests was simply not seen as being of benefit to the system.

As for transparency, the current publicising policy guideline was not seen as being problematic in our data. Overall, the moderate Finnish school evaluation culture was contrasted in the interviews either with the top PISA performers from South-East Asia or with the Anglo-American countries, while the recent QAE reforms in the Nordic countries were left unacknowledged. The arguments, which are used to justify the opposite publicising policy in the other Nordic countries – democracy, transparency, institutional accountability (or answerability) – were absent. Also, when discussing the parents' interests or role, the topic was touched on in a different way. Within the Finnish coordinative discourse, the parents were not only expected, but simultaneously normatively guided not to break 'the idea' of the Finnish school evaluation system, to follow a certain 'logic of appropriate behaviour' (March & Olsen 1989) by simply trusting the system and the autonomy of the schools.

In summary, we argued that the discussion concerning school evaluation practices and publicising policy in the Nordic region was formed in the interplay of three main discourses: *The global competence discourse* links the need for intensified evaluation measures and increased political accountability with the results of the international assessments in the context of economic competitiveness. However, whereas the lower PISA results in Sweden, Denmark and Norway justify the changes in the QAE practices, the Finnish comprehensive school is described as relatively competitive, despite the PISA decline, hence without similar pressure to reform. *A (neo)-liberal discourse* articulates the increased testing as a follow-up service for individual schools and pupils. However, within this discourse, the school-specific evaluation results reflect the fundamental principle of modern-day governance transparency but also stimulate the marketised environment of consumer behaviour and school choice. In contrast to the other Nordic countries, this discourse in Finland was most clearly absent.

On the contrary, the Finnish interviewees constructed their views rather on a *Nordic comprehensive school discourse*, which derives its legitimacy from the traditional idea of a common and equal Nordic comprehensive school, working as a counterforce against the neo-liberal discourse. Within this discourse, the Finnish policy guideline not to report the schools' performance at the school level is interpreted to serve society, to support the egalitarianism of the system, superior to individual or market logic interests.

As we concluded in the article, our interview data with Finnish key experts did not show features of new, reforming or challenging ideas about the current QAE policy guidelines. Thus, the decline in the Finnish PISA results did not mark a critical juncture for us, in which new discourses would start to originate or strengthen:

The change in the governance of education through increased accountability and transparency has not yet reached Finland. According to our analysis, these pressures simply do not resonate with the Finnish core beliefs on school evaluation and its benefits. Despite the gradual decline in more recent PISA assessments, the coordinative discourse produced by the Finnish elite has been consistent in setting the boundaries for public debate. Even though the Finnish school system has been subjected to critical observation several times in the news this year, neither the guidelines for a national testing system nor publicity issues have been seriously challenged on any front. The master discourse of Nordic egalitarianism, articulated through the autonomy of the Finnish teacher, trust in the system and the detriments of high-stakes testing, accompanied by the PISA success, has so far effectively controlled the faintest neo-liberal opinions. (Article III, p. 142.)

7 DISCUSSION AND THE KEY ARGUMENTS

In this research, I have scrutinised a topic of growing importance, namely the policy of publicising school performance indicators within the context of the Nordic countries. The research task and interest was built around a simple notion regarding the current publicising policies and practices within the Nordic region. While in Sweden, Denmark and Norway, the governments have decided to publicise a range of school performance indicators at the school level on official web portals, in Finland no similar information is provided. Thus, the aim of this research was to clarify how and why Finland has been able to resist the pressures of ‘global testing culture’ and the idea of publicising school-specific performance results.

The research consisted of three individual sub-studies, in which the institutionalisation of the publicising policy in Finland was contrasted with that in the other Nordic countries. Institutionalisation was understood in this research as historically evolved structures and practices, but above all, I was interested to scrutinise how the embedded rationalities and beliefs that underlie the publicising policy are shared in their discursive manifestations – in the policy argumentation and legitimisation, as well as in the policy problematisations. To capture these underlying rationalities, I focused on the formation of the coordinative policy discourse articulated by the key policy actors (see Schmidt 2008). The coordinative discourse reflects, but also shapes and defines the frames and the meanings for what is seen as ‘appropriate’ policy in society.

In the first sub-study (*Wallenius, T.J. (2015). Justifying opposite publication policies of school performance results in Finland and Sweden*), I analysed the argumentation on the publicising policy in Finland and Sweden in the 2000s. The primary data consisted of seven interviews with Finnish policy actors (collected in 2007–08), which I contrasted with the official policy texts of the Swedish National Agency for Education [Skolverket] and the Swedish FabQ interviews (Segerholm 2009). The analysis showed how the opposite publicising policies and their justifications became logically coherent when noticing ‘the image of the quality of the school system’ as a premise for the policy problematisation. These opposing images, furthered strongly by the PISA results in the 2000s, strengthened the legitimacy of the QAE policies in both countries. In addition, the analysis showed a clear difference in the policymaking rationality attached to publicising school performance. Whereas the Finnish rationality followed the idea of ‘quality through preventing inequality’, the Swedish rationality aimed to ‘raise the quality through competition’.

The second sub-study (*Wallenius T. (2016). National Testing of Pupils in Finland and Sweden in Light of Historical Institutionalisation [originally*

published in Finnish) contrasted the historical institutionalisation of the national level pupil testing practices in Finland and Sweden. This sub-study showed how the current policies in both countries can be seen to entail historically institutionalised path-dependent elements. The analysis showed how even in its early stages, the aims and the practices of national testing evolved along different trajectories, shaping the policy-making context of the 1990s in which the current policy guidelines of the comprehensive school QAE and publicising policy were outlined. The different historical experiences on the use and the benefits of standardised testing can be seen in both countries' current policies – for example, in the obligatory national testing in Sweden, but above all in the Finnish sample-based pupil testing, which prevents in practice the possibility of publicising comparable performance data between individual schools.

In the third sub-study (Wallenius, T., Juvonen, S., Hansen, P. & Varjo, J. (2018). *Schools, accountability and transparency - Approaching the Nordic school evaluation practices through discursive institutionalism*), the scope of contrasting was extended to cover Denmark and Norway. Here, the analysis consisted of 58 interviews with key policy actors, collected during 2015–17. There was special interest in the Finnish interviews, because the interviews were held after a decline in the Finnish PISA results, understood as a potential juncture for new ideas and discourses to emerge. However, the analysis showed how the Finnish discourse on the publicising policy remained different, not only from the Swedish, but also from the Danish and the Norwegian policy discourses, despite the acknowledged decline in the recent PISA results. The two essential concepts, accountability and transparency that in the other countries promoted the need for intensified QAE measures and the publicising of school-specific performance indicators, were not simply seen to benefit the Finnish school system nor the society by the Finnish informants.

Together, the three research articles showed that the deep-seated rationalities institutionalised in the policies, practices and policy discourses, strengthened further by Finland's initial PISA success, have provided a suitable platform for Finnish policy actors to control the coordinative policy discourse on the comprehensive school QAE policy and to resist effectively the pressures to publicise school-specific performance indicators. Thus, based on the findings of the three research articles and this Summary, I argue for acceptance of the following key arguments within my dissertation:

First, I argue that the decisions that led to the Finnish publicising policy entail path-dependent elements dating from before the 1990s and the emergence of the 'global testing culture' and school ranking policy. The historical analysis, in which the evolution of the national level pupil testing in Finland was contrasted with the situation in Sweden, showed clear differences in the rationalities concerning the benefits attached to the use of standardised pupil assessment for enhancing educational equality. Thus, when the question about publicising policy

came to be considered and decided on during the 1990s, the countries had already been directed to significantly different policy trajectories in their national level QAE policy.

Secondly, I argue that in contrast to the other Nordic countries, the publicising policy in Finland has taken a significant central role in comprehensive school QAE policymaking. The Finnish principle, not to publicise school-specific performance data became a key issue as the framework for the Finnish QAE was formulated in the 1990s, eventually defining even the method for the current national testing to be conducted by a sample-based method. In the other Nordic countries, the publicising policy did not receive similar emphasis in the policymaking. For example, in Sweden, the question was treated as a common principle of governance publicity and in Norway, despite the intense policy debate on its detrimental consequences, publicising school-specific data did not turn into an ultimate and rejective issue in the QAE policymaking.

Thirdly, in addition to the historically institutionalised elements, I argue that the capacity of the Finnish policy actors to control the discourse on the publicising policy differs from that in the other Nordic countries. This notion was strengthened throughout the research but was most clearly illustrated within Article III. Despite the decline in the Finnish PISA results, the current publicising policy was not problematised in our interview data. On the contrary, the Finnish coordinative policy discourse, which we described as a the Nordic comprehensive school discourse in Article III, but in this Summary I suggest to be named as the *depoliticisive discourse of school performance*, continued to set the limits for the ‘appropriate’ (March & Olsen 1989) policy and behaviour.

In my view, this *depoliticisive discourse of school performance* in the context of Finnish comprehensive school QAE policy is buffering, stagnating and normative by its nature. *Buffering*, since it highlights the negative consequences attached to a school-specific publicising policy; *stagnating*, since it muffles the need to reform the current policy, and *normative*, since it normatively guides the citizens to trust the Finnish comprehensive school system and ‘prevents’ the citizens from seeing themselves as eligible users of school-specific performance data.

Within the Finnish QAE policy, the *depoliticisive discourse of school performance* manages to exert effective control over the manifestations of the two central concepts, accountability and transparency, through which the school-specific publicising policy is typically being promoted. In Finland, the test-based accountability has a strongly negative connotation and any action or discursive opening proposing a change to the QAE guidelines may be interpreted as a challenge to the autonomy of Finnish schools and the teachers. In my data, in order to strengthen this discourse rhetorically, the Finnish QAE policy was often framed in the interviews against the ‘high accountability cultures’ (see e.g. OECD 2013, 64), without acknowledging greater variance in its forms. On the contrary, the

informants in Sweden, Denmark and Norway admitted an intentional increase in the accountability measures in their recent QAE policy. Still, the systems were described as being different from the high accountability cultures and the concept was discussed in practical terms, entailing a promise to monitor and ensure the quality of the school system.

The manifestations of transparency are also controlled carefully within this discourse, yet by a different technique, namely by depoliticising both the need and the right of the citizens to have access to comparable school performance data. Even if transparency is generally highly valued in Finnish governance – as it is in all the Nordic countries – in this particular field of school performance, the Finnish discourse differed from the other Nordic countries. Unlike the Swedish, Norwegian and the Danish interviewees, the Finnish informants did not discuss the publicising policy as an issue dealing either with the citizens' democratic rights or as an action supporting equal access to official information. Instead, this type of neo-liberal discourse was hardly evident in my data.

In my view, this silence or absence can be understood in two ways. On the one hand, we may think of it as reflecting an established way to perceive the question of publicising school performance in Finnish society. By this, I mean its taken-for-granted nature, and that the issue is simply not even acknowledged to be complex or problematised within Finnish society. However, depoliticising transparency may also be understood as strategic behaviour in controlling the coordinative discourse and its message. In cases when the attention on the differences of individual schools' performance was increased, the current Finnish QAE policy would appear problematic in terms of both transparency and democracy. Thus, to maintain the legitimacy of the current policy, depoliticising its connection to these in general highly valued principles of governance is nothing but strategically rational.

8 CONCLUDING REMARKS

I would now like to make some concluding remarks on the research topic and its findings. First, I want to comment on the research outcomes from a more theoretical point of view. Within neo-institutionalist policy research, the historical or the sociological approaches have typically been more prominent in explaining the sustainability of institutionalised practices, while the research on discursive institutionalism has focused on the mechanisms of change. Indeed, policy change comes subsequent to the emergence of new ideas and policy discourses.

However, this research has shown that when researching a policy topic that seems more or less unanimous or taken for granted, it is important to scrutinise the institutionalisation of the discursive elements that buffer the change. By this, I mean how the emerging ideas and discourses relate and challenge the earlier rationalities attached to the topic. The Finnish case shows this illustratively. For example, once the autonomy of the Finnish schools and the teachers has been interpreted as one key factor in explaining the Finnish PISA success, it becomes more difficult to counter-argue for its dismantling, despite the decline in later assessments. Thus, also the discourses may entail path-dependent elements that explain institutional sustainability rather than change.

In addition, this research has endorsed how the connotation of the two chameleon-like concepts in present-day governance, accountability and transparency, are not universal but without doubt are policy sector specific. In Finland, the boundaries of institutional trust in the core institutions has been tested on several occasions during the very recent years¹⁹. Even though we may treat these cases as being exceptional and single in their nature, the demands to increase institutional accountability and transparency have been seen as obvious and necessary actions in these policy sectors. In this sense, the Finnish comprehensive school QAE system and the publicising policy may add new points of view to the theoretical discussion on the mechanisms of accountability and transparency in governance. For example, if transparency is generally considered to further institutional trust in society, under what premises or circumstances would an institution or society actually benefit from a reverse action? In other words, what would the impact be of non-transparency? To scrutinise the interplay between

¹⁹ For example, in January 2019, the Finnish elderly care system confronted severe critique after shocking news on profit-making companies that run their businesses with minimal human resources and below any tolerable standards in the quality of care. In 2013, a ‘Hollywood-crime-story’ case of a respected police detective, who was eventually sentenced to imprisonment after having turned out being a mastermind in a massive drug trafficking business, raised the leadership and the practices of internal control of the police force in question. In 2019, scandalous behaviour of a high-ranking officer in a reserve army training resulted with several legal proceedings in the Finnish defense forces.

transparency, accountability, institutional trust and performance and their relationship to politicisation and depoliticisation would surely be an interesting research topic to conceptualise further.

My second point is more topic related and touches on future scenarios. The experiences from the other Nordic countries within this research have shown how the governments and the responsible school evaluation institutions have taken an active role in defining the ways to publicise school performance data. The technology has allowed new opportunities for data management and for its visual illustration. The content of the web portals has also been revised many times, which in my view shows that the governments are constantly seeking an optimal way to publicise the data. However, the aims, to combine both the idea of equal access to school performance and to control the negative consequences of school segregation, is complex and questionable. In some countries, the publicising of school performance has turned into for-profit business²⁰. At least from the Finnish perspective, this sounds highly unwarranted and unacceptable.

In this research, I have tried to avoid taking a normative stand about whether the performance indicators ought to be publicised school-specifically or not. On the contrary, my aim has rather been to increase the understanding on the elements in which both opposite policy solutions have become rationalised and institutionalised as legitimate social practices within the Nordic region. If any policy recommendations are to be given, I suggest that the Finnish policy actors keep observing the experiences and social consequences that follow from the current publicising policies in the other Nordic countries. However, the Nordic countries are the closest reference countries in Finnish comprehensive school policymaking. The experiences from Norway may especially turn out to be valuable since of all the Nordic countries, it is perhaps the Norwegian environment that most closely resembles the Finnish school system in its structures (e.g. mainly public schools, school choice taking place practically in only few bigger cities). What effects the official web portals have on the schools or more broadly in society is a topic to follow up attentively.

Thirdly, and finally, despite its taken-for-granted nature at this point, the societal environment of policymaking is constantly changing. So far, despite the changing power relations and various government compositions, the guidelines of the Finnish basic education QAE policy have remained the same and unchallenged for the last 20 years. The official definition has continuously emphasised that the aim of the Finnish comprehensive school QAE policy is for developmental purposes. Ironically, the principles behind the policy guidelines have left little leeway for the national level QAE system itself to develop.

²⁰ For example in the UK, the government publicises performance data from all educational levels (Gov.UK n.d.) but similar data with optional 'premium features' (better comparability, full data access etc.) is offered to the parents at cost by private companies (e.g. SchoolGuide n.d.).

Even though no official school-specific performance data are provided in Finland, the schools are a typical topic of discussion among the pupils' parents. In that light, it is interesting how the Finnish policy and its main discourse has managed to depoliticise the whole idea of having any official, publicly available data needed. One could ask with fairly good reason, if this is not a question of enhancing the families' equality, providing the families an equal access to information? One could ask whether the current policy fails to emphasise the role of unofficial information, in other words 'hot' knowledge (Ball & Vincent 1998) that we know that is unevenly distributed in the social networks across different social-class groups.

I am not saying that the Finnish policy should be reformed. We are aware of the detrimental effects that the school-specific publicising policy could entail. Still, the younger generations have already learnt that the information affecting their lives, whether on public or private institutions, is available online and is examinable. What I am saying is that this change in the mindset may start conflicting with the current publicising policy in Finland at some stage. If the 'image of the quality' of the Finnish comprehensive school system, referring here mostly to the PISA scores, continues to decline and diversify, more critical voices may arise, especially if pupils' achievements improve in the other Nordic countries at the same time. In this potential, if unwanted future scenario, the Finnish QAE model and its guidelines may start to appear stagnant and nonprogressive in contrast with the other Nordic countries. Then, at the latest, the sustainability of the core beliefs behind the Finnish QAE policy may be tested for real.

EPILOGUE

On 26 June 2020, the National Agency for Education in Sweden (Skolverket) came out with an announcement on fundamental changes in the policy of publicising school performance indicators in Sweden [translation by the author]:

Changed privacy policy affects access to certain statistics

From 1 September 2020, the National Agency for Education will only publish statistics at the national level. We are implementing this change because of Statistics Sweden's decision on a revised privacy policy, which means that information on independent schools must conform with policies on privacy. The matter has been dealt with by the Court of Appeal and the judgement has since gained legal force. Among other things, we have now been commissioned by the government to propose constitutional amendments to resolve the situation that has arisen.

In 2019, Statistics Sweden changed its interpretation of what counts as classified information. In a decision in December 2019, the Administrative Court of Appeal in Gothenburg ruled that information from Statistics Sweden on the throughput, grading and composition of students at independent schools is covered by privacy protocols in accordance with Section 8 of the 'Public Access to Information and Secrecy Act' for the protection of the individual principal. Since leave to appeal the decision has not been granted by the Supreme Administrative Court, the National Agency for Education can no longer publish statistics on independent schools. Since equal conditions will apply to municipal and independent schools, statistics on municipal schools will not be available either. The same applies to preschools and leisure centres.

'We are now working to find a solution so that parents, decision-makers within the school, school principals, researchers and other users can get collective information about the Swedish school system again', says department head Eva Durhán.

National Agency for Education in Sweden, 26 June 2020 (Skolverket, 2020.)

The abovementioned announcement from Sweden shows that the guidelines relating to national QAE policies and the use of school performance indicators in the Nordic countries are constantly open for change even in most unexpected ways, despite their institutionalised and established policies, practices and rationalities. The new policy guideline concerns a range of key statistics, including

information on the number of pupils, grades, the proportion of qualified teachers, among other comparative figures. How the solution will turn out in Sweden and what its implications will be in the other Nordic countries is an extremely interesting question to follow. The research topic will definitely provide intriguing aspects for researchers in the field of comparative policy research in education in the future.

REFERENCES

- Alvesson, M. & Sköldbberg, K. (2009). *Reflexive Methodology: New Vistas for Qualitative Research* (2nd ed.). London: Sage.
- Antikainen, A. (2006). In Search of the Nordic Model in Education. *Scandinavian Journal of Educational Research*, 50(3), 229–243.
- Aurén, H. & Joshi, D. (2016). Teaching the World That Less Is More: Global Education Testing and the Finnish National Brand. In W. C. Smith (Ed.) *The Global Testing Culture: Shaping Education Policy, Perceptions, and Practice*, 63–83. Oxford: Symposium.
- Bacchi, C. (2009). Analysing policy: *What's the problem represented to be?* Frenchs Forest: Pearson Education.
- Ball, C. (2009). What is Transparency? *Public Integrity*, 11(4), 293–307. DOI: 10.2753/PIN1099-9922110400.
- Ball, S., & Vincent, C. (1998). 'I Heard It on the Grapevine': 'Hot' Knowledge and School Choice. *British Journal of Sociology of Education*, 19(3), 377–400.
- Bauhr, M. & Grimes, M. (2012.) *What is Government Transparency?: New Measures and Relevance for Quality of Government*. University of Gothenburg, QoG Working Paper Series 2012:16.
- Berman S. (1998). *The Social Democratic Movement: Ideas and Politics in the Making of Interwar Europe*. Cambridge, MA: Harvard University Press.
- Beuchert, L. V. & Nandrup, A. B. (2015). *The Danish National Tests – A Practical Guide*. Aarhus University: Department of Economics and Business, Economics Working Papers 2014:25.
- Beveridge, S. (1992). 'This is your charter...': parents as consumers or partners in the educational process? *Early Years*, 13(1), 12–15. DOI: 10.1080/0957514920130103.
- Blossing, U., Imsen, G., & Moos, L. (Eds.). (2014). *The Nordic Education Model. 'A School for All' Encounters Neo-Liberal Policy*. Dordrecht: Springer.
- Bowerman, M., Raby, H. & Humphrey, C. (2000). In Search of the Audit Society: Some Evidence from Health Care, Police and Schools. *International Journal of Auditing*, 4(1): 71–100. DOI: 10.1111/1099-1123.00304.
- Bracci, E. (2009). Autonomy, responsibility and accountability in the Italian school system. *Critical Perspectives on Accounting*, 20(3), 293–312. DOI: 10.1016/j.cpa.2008.09.001.
- Broadfoot, P. (2000). Preface. In A. Filer (Ed.) *Assessment: Social Practice and Social Product*, (pp. ix–xii). London and New York: RoutledgeFalmer.
- Burgess, S., Wilson, D. & Worth, J. (2010). *A natural experiment in school accountability: the impact of school performance information on pupil*

- progress and sorting*. CMPO Working Paper Series No. 10/246, University of Bristol.
- Capoccia, G. (2015). Critical junctures and institutional change. In J. Mahoney & K. Thelen (Eds.), *Advances in Comparative-Historical Analysis (Strategies for Social Inquiry)*, pp. 147–179. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781316273104.007.
- Capoccia, G. & Kelemen, R. D. (2007). The Study of Critical Junctures: Theory, Narrative, and Counterfactuals in Historical Institutionalism. *World Politics*, 59(3), pp. 341–369.
- Carstensen, M. B. & Schmidt, V. A. (2016). Power through, over and in ideas: conceptualizing ideational power in discursive institutionalism. *Journal of European Public Policy*, 23(3), 318–337. DOI: 10.1080/13501763.2015.1115534.
- Connolly, W. E. (1983). *The Terms of Political Discourse*. 2nd ed. Princeton, NJ: Princeton University Press.
- Dahler-Larsen, P. (2011). *The Evaluation Society*. California: Stanford University Press.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106(6), 1047–1085.
- Department for Education. (1994). *Our children's education: the updated parent's charter*. London, DfE.
- Department of Education and Science. (1991). *Schools to publish all examination results this summer*. DES News, 185/91.
- DiMaggio, P. J. & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48(2), pp. 147–160.
- Dubnick, M. J. (2014). Accountability as a Cultural Keyword. In M. Bovens, R. E. Goodin & T. Schillemans (Eds.) *The Oxford Handbook of Public Accountability*. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199641253.001.0001.
- Elstad, E. (2009). Schools which are named, shamed and blamed by the media: school accountability in Norway. *Educ Asse Eval Acc* 21, 173–189. DOI 10.1007/s11092-009-9076-0.
- Erkkilä, T. (2010). *Reinventing Nordic Openness: Transparency and State Information in Finland*. Helsinki: University of Helsinki, Department of Political and Economic Studies.
- Erkkilä, T. (2012). *Government Transparency: Impacts and Unintended Consequences*. Basingstoke: Palgrave.
- Esping-Andersen, G. (1990). *The Three Worlds of Welfare Capitalism*. Princeton, New Jersey: Princeton University Press.
- Eurydice (2009). *National testing of pupils in Europe: Objectives, organization and use of results*. EACEA: Eurydice. Bryssel. DOI: 10.2797/18294.

- Fox, J. (2007). The uncertain relationship between transparency and accountability. *Development in Practice*, 17(4-5), 663–671. DOI: 10.1080/09614520701469955.
- Friskolornas riksförbund. (2017.) *Sveriges bästa och sämsta grundskolor 2017. Skolenheter som presterar bäst och sämst slutbetyg i årskurs 9 med hänsyn till elevsammansättning enligt Skolverkets SALSA-statistik* [Sweden's best and worst primary schools 2017. School units that perform best and worst by final grade in year 9 with regard to student composition according to the National Agency for Education's SALSA statistics]. <https://www.friskola.se/2017/11/03/sveriges-basta-och-samsta-grundskolor-2017/>, 3.11.2017.
- Goldstein, H. & Spiegelhalter, D., J. (1996). League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 159(3), 385–443.
- Green, A. 1997. *Education, globalization and the nation state*. New York: St. Martin's Press.
- Grimmelikhuijsen, S., Porumbescu, G., Hong, B. & Im, T. (2013). The Effect of Transparency on Trust in Government: A Cross-National Comparative Experiment. *Public Administration Review*, 73(4), 575–586. DOI: 10.1111/puar.12047.
- Hall P. (1993). Policy paradigms, social learning and the state: the case of economic policymaking in Britain. *Comp. Polit.*, 25, 275–296.
- Hall, P. A. & Taylor, R. C. R. (1996). Political science and the three new institutionalisms. *Political Studies*, 44(5), 936–957.
- Hallgarten, J. (2001). School league tables: Have they outlived their usefulness? *New Economy*, 8(4), 189–196.
- Hansen, H. P., Wallenius, T., Juvonen, S., & Varjo, J. T. (2019). Moving landscapes of Nordic basic education: Approaching shifting international influences through narratives of educational experts. *Compare*, 1-19. <https://doi.org/10.1080/03057925.2018.1557509>
- Hanushek, E. A. & Raymond, M. E. (2005). Does School Accountability Lead to Improved Student Performance? *Journal of Policy Analysis and Management*, 24(2), 297–327.
- Hay, C. (2006). Constructivist Institutionalism. In R.A.W. Rhodes, S.A. Binder & B.A. Rockman (Eds.) *The Oxford Handbook of Political Institutions*. Oxford: Oxford University Press, pp. 56–74.
- Hood, C. (2010). Accountability and Transparency: Siamese Twins, Matching Parts, Awkward Couple? *West European Politics*, 33(5), 989–1009. DOI: 10.1080/01402382.2010.486122.
- Hovdhaugen, E., Vibe, N., & Seland, I. (2017). National test results: Representation and misrepresentation. Challenges for municipal and

- local school administration in Norway. *Nordic Journal of Studies in Educational Policy*, 3(1), 95–105.
- Hudson, C. (2007). Governing the Governance of Education: The State Strikes Back? *European Educational Research Journal*, 6(3), 266–282.
<https://doi.org/10.2304/eeerj.2007.6.3.266>
- Jakku-Sihvonen, R. (2013). Oppimistulosten arviointijärjestelmistä ja niiden kehittämishaasteista [On Pupil Assessment Systems and Their Developmental Challenges]. In A. Räisänen (Ed.) *Oppimisen arvioinnin kontekstit ja käytännöt*. Opetushallituksen raportit ja selvitykset 2013:3, 13–36. [The Contexts and Practices of Pupil Assessment. Reports of the NBoE 2013:3.] Helsinki: Opetushallitus.
- Jones, G. W. (1992). The Search for Local Accountability. In S. Leach (Ed.) *Strengthening Local Government in the 1990s*. London: Longman, pp. 49–78.
- Karsten, S., Visscher, A. & de Jong, T. (2001). Another side to the coin: the unintended effects of the publication of school performance data in England and France. *Comparative Education*, 37(2), 231–242.
- Kauko, J., Centeno, V., Piattoeva, N., Hinke Dobrochinski Candido, H., Gurova, G., Medvedeva, A., Santos, Í., Suominen, O. & Zhou, X. (2018). Layers of reflectivity in comparative research. In J. Kauko, R. Rinne & T. Takala (Eds.) *Politics of Quality in Education: A Comparative Study of Brazil, China, and Russia*. London: Routledge.
- Kauko, J., Rinne, R. & Takala, T. (Eds.) (2018). *Politics of Quality in Education: A Comparative Study of Brazil, China, and Russia*. London: Routledge.
- Kellaghan, T., Stufflebeam, D. & Wingate, L. (2003). Introduction. In T. Kellaghan & D. Stufflebeam (Eds.) *International Handbook of Educational Evaluation* (pp. 1–6). Dordrecht: Kluwer Academic Publishers.
- Konttinen, R. (1995). Arvostelusta näyttöön. Koulutuksen arvioinnin kehityspiirteitä Suomessa [From Assessment to Evaluation. Development of Educational Evaluation in Finland]. *Kasvatus*, 26(1), 6–14.
- Kouvo, A. (2011). The sources of generalized trust and institutional confidence in Europe. *Research on Finnish Society*, Vol. 4 (2011), pp. 29–40.
- Landman, T. (2008). *Issues and Methods in Comparative Politics: An Introduction*. 3rd Ed. London and New York: Routledge.
- Laukkanen, R. (1998). *Opetustoimen evaluaatioajattelun kehitys Suomessa 1970-luvulta 1990-luvulle* [Development of Conceptions of Evaluation in Central Educational Administration in Finland from the 1970s to the 1990s]. Jyväskylän yliopisto. Koulutuksen tutkimuslaitos. Tutkimuksia 5.
- Law 628/1998. *Basic Education Act*.

- Lecours, A. (2005). New Institutionalism: Issues and Questions. In A. Lecours (Ed.) *New Institutionalism: Theory and Analysis* (pp. 3–26). Toronto; Buffalo; London: University of Toronto Press.
- Lindblad, S. & Popkewitz, T. (Eds.) (1999). *Education governance and social integration and exclusion: National cases of educational systems and recent reforms*. Uppsala reports on education 34.
- Lingard, B. (2010). Policy borrowing, policy learning: Testing times in Australian schooling. *Critical Studies in Education*, 51(2), 129–147.
- Listhaug, O. & Ringdal, K. (2007). *Trust in Political Institutions: The Nordic Countries Compared with Europe*. A Conference paper for the Norwegian Political Science Meeting, NTNU, Trondheim, Jan 3–5, 2007.
- Ljung, B.-O. (2000). *Standardproven – 53 år i skolans tjänst*. Rapport från PRIM-gruppen nr. 17 [The Standardtest – 53 years in the school's service. Report 17 of the PRIM-group]. Stockholm: Lärarhögskolan i Stockholm, Institutionen för undervisningsprocesser, kommunikation och lärande.
- Lundahl, C. (2009). *Varför nationella prov? – framväxt, dilemma, möjligheter* [Why national tests? Origin, Dilemma, Opportunities]. Lund: Studentlitteratur.
- Lundahl, L., Erixon Arreman, I., Holm, A.-S. & Lundström, U. (2013). Educational marketization the Swedish way. *Education Inquiry*, 4(3), 497–517. DOI: 10.3402/edui.v4i3.22620.
- Lundgren, U.P. (1990). Educational Policymaking, Decentralization and Evaluation. In M. Granheim, M. Kogan & U.P. Lundgren (Eds.) *Evaluation as Policymaking: introducing evaluation into a national decentralized educational system*. London: Jessica Kingsley.
- Madaus, G. F., Russell, M. K. & Higgins, J. (2009). *The Paradoxes of High Stakes Testing: How They Affect Students, Their Parents, Teachers, Principals, Schools, and Society*. Charlotte, N.C.: Information Age Publishing.
- Mahoney, J. (2000). Path dependence in historical sociology. *Theory and Society*, 29, 507–548.
- March, J. & Olsen, J. (1989). *Rediscovering institutions: The organizational basis of politics*. New York: Free Press.
- Marien, S. (2011). Measuring Political Trust Across Time and Space. In M. Hooghe & S. Zmerli (Eds.) *Political Trust. Why Context Matters*, (pp. 13–46). Colchester: ECPR Press.
- Maw, J. (1999). League tables and the press – value added? *Curriculum Journal*, 10(1), 3–10.
- Meyer, H.-D. & Benavot, A. (Eds.) (2013). *PISA, Power, and Policy: the emergence of global educational governance*. Oxford: Symposium books.

- Meyer, J. W. & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, 83, pp. 340–363.
- Ministry of Children and Education. (2017). *Retningslinjer for brug af data fra de nationale test til statistiske og videnskabelige undersøgelser* [Guidelines for the use of data from national tests for statistical and scientific studies]. Undervisningsministeriet. ISBN: 978-87-603-3075-9.
- Mulgan, R. (2000). 'Accountability': An Ever-Expanding Concept? *Public Administration*, 78(3), 555–573.
- Mykkänen, J. (2001). Eliittihaastattelu [Elite Interview]. *Politiikka*, 43(2), 108–127.
- National Board of Education. (1995). *Koulutuksen tuloksellisuuden arviointimalli* [A Framework for Evaluating Educational Outcomes in Finland]. Helsinki: Opetushallitus (Arviointi 9/1995).
- National Board of Education. (1998a). *Kansallinen oppimistulosten arviointijärjestelmä* [National Level Pupil Testing System]. Helsinki: Opetushallitus (Arviointi 4/1998).
- National Board of Education. (1998b). *Koulutuksen tuloksellisuuden arviointimalli* [A Framework for Evaluating Educational Outcomes in Finland]. Helsinki: Opetushallitus (Arviointi 7/1998).
- Niukko, S. (2006). 'Yhteistyötä ilman riskejä'? *OECD:n rooli Suomen koulutuspolitiikassa* [Co-Operation without Risks? The Role of the OECD in Finnish Education Policy]. Turun yliopiston julkaisuja. Sarja C. Osa 251.
- Nordic Council of Ministers. (2018). *Northern Lights on TIMSS and PISA 2018*. TemaNord 2018:524.
- Norwegian Directorate for Education and Training. (2011). *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Norway*. <http://www.oecd.org/education/school/47088605.pdf>
- Noy, C. (2008). Sampling knowledge: The hermeneutics of snowball sampling in qualitative research. *International Journal of Social Research Methodology*, 11(4), 327–344.
- OECD. (2013). *Synergies for Better Learning. An International Perspective on Evaluation and Assessment*. OECD Reviews of Evaluation and Assessment in Education.
- Ouakrim-Soivio, N. (2013). *Toimivatko päättöarvioinnin kriteerit? Oppilaiden saamat arvosanat ja Opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiserojen mittareina* [Are the criteria for final grading functioning? Student grades and NBE reviews on learning outcomes as indicators of between-school differences]. Raportit ja selvitykset 2013:9. Helsinki: Opetushallitus.

- Ozga, J. & Jones, R. (2006). Travelling and embedded policy: the case of knowledge transfer. *Journal of Education Policy*, 21(1), 1–17. DOI: 10.1080/02680930500391462.
- Perryman, J., Ball, S., Maguire, M. & Braun, A. (2011). Life in the Pressure Cooker – School League Tables and English and Mathematics Teachers’ Responses to Accountability in a Results-Driven Era. *British Journal of Educational Studies*, 59(2), 179–195. DOI: 10.1080/00071005.2011.578568.
- van Petegem, P., Vanhoof, J., Daems, F. & Mahieu, P. (2005). Publishing information on individual schools? *Educational Research and Evaluation*, 11(1), 45–60.
- Pierson, P. (2000). The limits of design: Explaining institutional origins and change. *Governance*, 13(4), 475–499.
- Pitkänen, H. (2019). *Arviointi, tieto ja hallinta: Peruskoulun paikallisen arvioinnin genealogia* [Evaluation, Knowledge and Power. Genealogy of local quality evaluation in the field of comprehensive education]. Helsinki: Helsingin yliopisto.
- Powell, W. W. & DiMaggio, P. J. (1991). *The New Institutionalism in Organizational Analysis*. Chicago and London: University of Chicago Press.
- Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Power, S. & Frandji, D. (2009). Education markets, the new politics of recognition and the increasing fatalism towards inequality. *Journal of Education Policy*, 25(3), 385–396.
- Rizvi, F. & Lingard, B. (2010). *Globalizing Education Policy*. London: Routledge.
- Rosenkvist, M. (2010). *Using student test results for accountability and improvement: A literature review*. OECD Education Working Papers. Paris: OECD.
- Ruusuvuori, J. & Tiittula, L. (2005). Johdanto [Introduction]. In J. Ruusuvuori & L. Tiittula (Eds.) *Haastattelu. Tutkimus, tilanteet ja vuorovaikutus* [Interview. Research, Context and Interaction]. Tampere: Vastapaino.
- Räisänen, A. (Ed.) (2013). *Oppimisen arvioinnin kontekstit ja käytännöt. Raportit ja selvitykset 2013:3* [The Contexts and Practices of Pupil Assessment. Reports of the NBoE 2013:3]. Helsinki: Opetushallitus.
- Saari, H. (1983). Koulukoetoinnin periaatteista peruskoulussa [On Principles in Pupil Testing]. In V. Hirvi (Ed.) *Peruskoulun kehittäminen tutkimustulosten perusteella. Tutkijoiden artikkeleita peruskoulusta ja sen kehittämisestä* [Developing the Comprehensive School by Research]. Jyväskylän yliopisto. Kasvatustieteiden tutkimuslaitos. Selosteita ja tiedotteita 209/1983, 104–5.

- Sahlberg, P. (2011). *Finnish Lessons: What Can the World Learn from Educational Change in Finland?* New York: Teachers College Press.
- Schmidt, V. A. (2008). Discursive institutionalism: The explanatory power of ideas and discourse. *Annual Review of Political Science*, 11, 303–326.
- Schmidt, V. A. (2010). Taking ideas and discourse seriously: Explaining change through discursive institutionalism as the fourth ‘new institutionalism’. *European Political Science Review*, 2(1), 1–25.
- Shiel, G., Kellaghan, T. & Moran, G. (2010). *Standardised testing in lower secondary education*. Dublin: National Council for Curriculum and Assessment, Research Report no. 12.
- Simola, H. (2005). Koulukohtaiset oppimistulokset ja julkisuus [School-Specific Learning Results and Publicity]. *Yhteiskuntapolitiikka*, 70(2), 179–187.
- Simola, H., Kauko, J., Varjo, J., Kalalahti, M. & Sahlström, F. (2017). *Dynamics in Education Politics – Understanding and Explaining the Finnish Case*. London: Routledge.
- Simola, H., Rinne, R., Varjo, J., Pitkänen, H. & Kauko, J. (2009). Quality assurance and evaluation (QAE) in Finnish compulsory schooling: a national model or just unintended effects of radical decentralisation? *Journal of Education Policy*, 24(2), 163–178.
- Simola, H., Varjo, J. & Rinne, R. (2010). Vasten valtavirtaa – kontingenssi, polkuriippuvuus ja konvergenssi suomalaisen perusopetuksen laadunpariointimallin kehityskuluissa [Against the mainstream – contingency, path dependence and convergence of quality evaluation models in the Finnish basic education system]. *Hallinnon tutkimus*, 29(4), 285–302.
- Sinclair, A. (1995). The Chameleon of Accountability: Forms and Discourses. *Accounting, Organizations and Society*, 20(2), 219–237.
- Skedsmo, G. (2011). Formulation and realisation of evaluation policy: inconcistencies and problematic issues. *Educ Asse Eval Acc* 23, 5–20. DOI: 10.1007/s11092-010-9110-2.
- Smith, W.C. (Ed.) (2016). *The Global Testing Culture: shaping education policy, perceptions, and practice*. Oxford: Symposium books.
- Steiner-Khamsi, G. (Ed.). (2004). *The Global Politics of Educational Borrowing and Lending*. New York and London: Teachers College Press.
- The Finnish Advisory Board on Research Integrity. (2012). *Responsible conduct of research and procedures for handling allegations of misconduct in Finland*. https://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf
- Thelen, K. (1999). Historical institutionalism in comparative politics. *Annu. Rev. Polit. Sci.*, 2, 369–404.
- Telhaug, A. O., Mediås, O. A. & Aasen, P. (2006.) The Nordic model in education: Education as part of the political system in the last 50 years. *Scandinavian Journal of Educational Research*, 50(3), 245–283.

- Toulmin, S. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.
- Tremblay, M.-A. (1982). The key informant technique: A non-ethnographic application. In R. Burgess (Ed.), *Field research: A sourcebook and field manual* (pp. 151–161). London: Routledge.
- Varjo, J., Simola, H. & Rinne, R. (2016). *Arvioida ja hallita. Perään katsomisesta informaatio-ohjaukseen suomalaisessa koulupolitiikassa* [To Evaluate and Govern: From Looking After to Steering by Information in Finnish Education Policy]. Jyväskylä. Suomen Kasvatustieteellinen Seura. Kasvatusalan tutkimuksia – Research in Educational Sciences 70.
- Visscher, A. J. (2001). Public school performance indicators: Problems and recommendations. *Studies in Educational Evaluation*, 27(2001), 199–214.
- West, A. & Pennell, H. (2000). Publishing school examination results in England: incentives and consequences. *Educational Studies*, 26(4), 423–436.

List of www-site references:

- Gov.UK. (n.d.) Find and compare schools in England. Retrieved from <https://www.compare-school-performance.service.gov.uk/>, 7.1.2020.
- Ministry of Children and Education (n.d.). Skoletal web portal. Retrieved from https://uddannelsesstatistik.dk/Pages/grundskolen_overblik.aspx, 6.3.2019.
- SchoolGuide (n.d.) Retrieved from <https://www.schoolguide.co.uk/>, 7.1.2020.
- Skoleporten (n.d.). Retrieved from <https://skoleporten.udir.no/#> , 6.3.2019.
- Skolverket. (2020). *Förändrad sekretesspolicy påverkar tillgång till viss statistik*. [Changed privacy policy affects access to certain statistics]. Retrieved from <https://www.skolverket.se/skolutveckling/statistik/arkiverade-statistiknyheter/statistik/2020-06-26-forandrad-sekretesspolicy-paverkar-tillgang-till-viss-statistik> , 3.8.2020.
- Skolverketa. (n.d.). Skolverket's SIRIS (Skolverkets Internetbaserade Resultat- och kvalitets Informations System) web portal, retrieved from www.siris.skolverket.se
- Skolverketb. (n.d.). Skolverket's Välja skola web site. Retrieved from <http://valjaskola.se/index.html>, 6.3.2019. [Note: the address has been removed during 2019].
- Skolverketc. (n.d.). Skolverket's SALSA (Skolverkets Arbetsverktyg för Lokala SambandsAnalyser) web portal. Retrieved from <https://www.skolverket.se/skolutveckling/statistik/om-skolverkets->

statistik/salsa-statistisk-modell#h-SALSAskainteanvandasfor,

31.12.2019.

Søndagsavisen (2015). [See the whole list – The best and the worst schools in Denmark, 22.8.2015.] Retrieved from

<https://www.sondagsavisen.dk/familie-kaerlighed/2015-08-22-se-hele-listen-her-er-danmarks-bedste-og-vaerste-skole/>, 31.12.2019.

11 Justifying opposite publication policies of school performance results in Finland and Sweden

Introduction

The question whether to publish school-specific pupils' assessment results or not has been under puzzling discussion for more than last 20 years. The dichotomic question is said to have 'two sides of the coin' – on one hand, the supporters have stressed the accountability of public institutions and the need of information for parental choice. On the other hand, many negative and unintended side-effects have been brought out. Result-ordered listings, popular especially in the media, are said to give a simplified and misleading picture of the schools and accelerate the differentiation between 'good' and 'poor' performers. Either way, the culture of 'school ranking', originally characteristic for the Anglo-American educational governance, has been described to spread as an 'irreversible trend worldwide' in compulsory schooling (Karsten, Visscher, & de Jong, 2001).

This global trend has become evident in the Nordic countries where all the other countries but Finland have started (or are cur-

rently outlining) to publish more detailed performance results based on standardized national testing (Eurydice, 2009; Nusche, Earl, Maxwell, & Shewbridge, 2011). The development has been somewhat surprising, as the attempts to make school performance more visible and increasing competition between schools have been considered to be contradictory to the traditional core values of the Nordic comprehensive school system such as equality and school autonomy (Telhaug, Mediås, & Aasen, 2006; Antikainen, 2006; Rinne, 2000). Maybe most unexpected, the speed and the extent of market-based educational reforms have been most significant in Sweden, which used to represent the ‘model-country’ of the Nordic education system and welfare-state regime (Esping-Andersen, 1990).

In Finland the question of specified results has turned out to be very complex, even sensitive. Despite the general principle of transparency in governance, the Finnish policy actors have consistently responded very sceptically, even hostilely to any ‘school ranking lists’ or ‘league tables’ in basic education. So far, the resisting attitude has been dealt with a peculiar silent cultural consensus – the politicians, educational officials, teachers and principals, even most of the parents seem to oppose the idea as harmful for the Finnish school system (Rinne, Simola, Mäkinen-Streng, Silmäri-Salo, & Varjo, 2011; Välijärvi, 2012).

The aim of this article is to compare how the opposite policies on publishing school-specific performance data in Finland and Sweden have emerged and how they are justified. The research frame is by its nature retrospective, as it focuses on the justifications of political decisions made in the past. The actual empirical analysis consists of 7 interviews of Finnish policy actors collected in a previous FabQ research project (see chapter Interview data) in 2007. These results are then mirrored with the findings of the Swedish interviews (Segerholm, 2009) and other literature to construct the contextual frame in which the decisions have taken place. A new theoretical and methodological combination to critical policy analysis is provided as Carol Bacchi’s (2009) ‘what’s the problem represented to be’ approach is combined with Stephen E. Toulmin’s (1958)

Model of Argumentation. The article will be ended by a brief discussion with future predictions of the QAE practices in the Nordic context.

The two-sided quest for quality through the publication of the performance indicators

The publication of school-specific performance indicators by government's educational agencies started in the 1980s from the Anglo-American countries (UK, United States). At first this information contained e.g. average percentage rates of the pupils passing a certain criteria in the national standardized tests, school-leaving grades as well as background figures on funding or pupil-teacher ratio. Public evaluation results were argued to help parents to make 'informed school choices'. By increasing visible accountability it was also meant to improve the efficiency, especially of the public schools. During the 1990s the evaluation methods were heavily criticised. One central complaint was that the results did not observe any information of the learning progress in the schools. The critique led gradually in the implementation of the nowadays commonly used *value-added* measurements, which present the performance results in respect of pupils' previous grades, socio-economic backgrounds etc. (Karsten et al., 2001; Maw, 1999.)

Despite the attempt to develop more comprehensive or more ethical evaluation and publishing methods the scepticism has not vanished. Instead, as another side of the coin the publication is considered to have many negative consequences. Even if the results were seen as 'objective information', the rankings inevitably highlight the differences between schools and feed the idea of 'winners' and losers'. This has been the central point of the critics – that public evaluation results rather accelerate the differentiation between the schools and hence even lower the average quality. The publicity is also feared to affect the internal school-work or pupil-intake criteria as teachers and pupils are driven to perform successfully in the tests. (Allen & Burgess, 2011; Karsten et al., 2001; van Pe-

tegem, Vanhoof, Daems, & Mahieu, 2005; West & Pennell, 2000.)

After all, the publication of the school-specific performance results is tightly connected to the market-oriented logic and reforms in education. Instead of being just objective information, it is an essential element to create and stimulate market mechanisms in education (Power & Frandji, 2009, p. 386; West & Pennell, 2000). As part of the neo-liberal education policy and QAE activities the publication policy follows noticeably the logic of raising the overall quality of the schools by increasing the competition between individual schools more visibly (Friedman & Friedman, 1980; Chubb & Moe, 1990). Even though many empirical findings do not back the presumptions of the market-oriented advocates in education, the global appeal to markets has been strong among policymakers (Lauder & Hughes, 1999).

Two concepts in the Nordic context must be separately noted to demonstrate the complexity of the question. In the Nordic basic education *the educational equality* has been the most important value ever since the birth of the comprehensive school system in the 1960s/70s. The market-oriented reforms such as parental choice have later on challenged this central principle in all Nordic countries and at first sight the decision to publish school performance results could be counted among these by its segregative nature. However, the arguments for the publishing can also be grounded on equality. Without comprehensive information of the schools the choice is made partly on schools' reputation (Kosunen, 2014) and is connected to families' cultural capital, networks, resources etc., in other words 'silent information'. In this sense, if the information was made public for all, it might be argued also as an act for supporting equality.

The concept of *transparency* makes the question of publishing school performance data – especially for Finland – even more complex. On one hand, transparency has been a key principle in the Nordic governance and democracy – 'the Nordic openness'. Sweden represents in the 'principle of publicity' even globally an extreme case as the world's first law allowing citizens access to government documents was issued already in the 18th century. Even

though the Finnish Act on the Publicity of Official Documents (year 1951) was progressive in comparison to many countries, the autonomy era under tsarist rule in the 1900th century and the Cold War era with limited public debate are said to have left cultural features to the Finnish governance and publicity. (Erkkilä, 2010, p. 8–11.) However, as Erkkilä (2010) notes, the ever growing demands for transparency have been linked for the last decades closer with economic efficiency – thus, it might be expected that the pressure to publish also the compulsory school performance results would grow in Finland as well.

The differentiation of the Finnish and the Swedish QAE policies

In the turn of the 1990s both Finland and Sweden went through a heavy decentralisation and deregulation process in education policy and governance. Strengthening local autonomy in education was prioritized and evaluation arranged mainly by the new national agencies founded in 1991 (Finnish National Board of Education FNBE, Swedish National Agency for Education SNAE) replaced the former central-governed and very detailed norm-steered governance in compulsory schooling:

[In Finland] by the early 1990s all traditional forms of control over the teacher's work such as school inspections, a detailed national curriculum, officially approved teaching materials, weekly timetables based on the subjects taught and class diaries in which the teacher had to record what was taught each hour had been eliminated. The only remaining control mechanism is set minimum numbers of lessons to be taught in each subject in each school. The inspectorate, traditionally hated by teachers and municipalities, opposed the idea of local freedom. All these traditional means of control were to be replaced by evaluation, realised by the municipal and national authorities. (Simola, Rinne, Varjo, Pitkänen, & Kauko, 2009, p. 167.)

However, during the 1990s and by accelerating force in the 2000s, the evaluation practices in Finland and Sweden developed gradu-

ally in different directions. The Finnish compulsory schooling QAE policy in whole has been described as an ‘upstream policy’ based on four national features: 1) evaluation is above all for developing educational services and not an instrument of administrative control; 2) the information produced through evaluation serves the administrative bodies and the schools rather than the public or families; 3) no support for ranking lists or the publication of school-specific performance indicators, and 4) no intention for making schools or teachers accountable for learning results (Simola et al., 2009, p. 171–172). The question of publicity got a judicial nature as two separate appeals to the regional administrative courts made by the media tested the Finnish stand against educational league tables in 2000 and 2003. Characteristically, despite the Supreme Administrative Court’s order to hand over school-specific evaluation results in one of the cases, the pressure for the wider use of specified data in the media has remained low on compulsory schools – in the upper secondary schools the grade point averages and the matriculation exam results are on the contrary annually recognized in the tabloids (Simola, 2005a; Jakku-Sihvonen, 2010).

The Finnish evaluation culture in the basic education system has been contradictory to the global trend what Pasi Sahlberg (2011) has called ‘the Global Education Reform Movement’, GERM. Instead of emphasizing test-based accountability and control, the Finnish evaluation system has rested on shared responsibility and trust in the professionalism of the teachers and the principals and sample-based student assessments (Sahlberg, 2011, p. 103). This evaluation culture of relative high school autonomy with low accountability and testing has been many times highlighted also as one of the key factors behind the Finnish success in the OECDs PISA (The Programme for International Student Assessment) survey (e.g. Linnakylä & Välijärvi, 2005, p. 261–273).

Without underestimating the Finnish educational reforms in the 1990s (Varjo, 2007), in comparison to Sweden the extent of the reforms towards marketization, privatization and increased competition has been modest. For example, in Finland the school choice, ratified in the 1999 Basic Education Act (Law 628/1998) is still

made inside the public comprehensive school system. It is important to note, that in Finland the school choice was primarily argued to increase pupils' motivation, not to create competition between schools (Seppänen, 2006). In Sweden instead the idea of creating educational markets has been more influentially present. Even though the market-driven reforms may have been even more visible in the upper secondary level (Lundahl, 2012) the structures of the compulsory schooling have changed significantly. School choice was introduced in 1992, also the number of privately operated schools increased rapidly (Björklund, Clark, Edin, Fredriksson, & Krueger, 2005, p. 6; Bunar, 2010). As for QAE practices, the trend has been complying with the GERM. The national school inspections have become more and more structured and frequent, standardized national tests have been run compulsory for all pupils in Swedish, English and Mathematics since the middle of the 1990s by the SNAE, and finally, the school-specific reports of the inspections and the quantitative results of the assessments have been published for all for more than last 10 years (Segerholm, 2009).

The research questions, interview data and methodological choices

In this article I am interested primarily how the educational policy actors justify the practiced policy. In addition, the contextual frame behind the decision-making must be taken into account to explain the opposite policies. Here a view of critical policy analysis is taken. Even though we would not totally deny the conventional idea that policymaking is more or less rational and consequential to some real issues or experiments, critical policy analysis aims to shift the focus from the outcomes or the results of the policy to the contextual premises and the prerequisites behind the policy that are many times taken for granted or left implicit. A more detailed description of the methodology is presented below to answer the following research questions:

- 1) How are the opposite publishing policies on school-specific performance results justified in Finland and Sweden?
- 2) To what ‘problem’ the practiced policies are aimed to answer?

Interview data

As mentioned before, the nature of the research frame is retrospective. The actual analysis consists of 7 interviews of the Finnish educational policymakers and officials in highest position. The interviews were collected in 2007–8 as part of an international research project *Fabricating Quality in European Education* (see Ozga, Dahler-Larsen, Segerholm, & Simola, 2011 [notation: the writer has not personally participated in the FabQ project]). Having represented the central educational agencies (mostly FNBE), the interviewees are treated here as central policy actors for drawing the national guidelines of the compulsory schooling QAE policy – understandably their views aim to legitimate the national policy rather than its criticism. The interview citations are presented anonymously following the research ethics agreed in the FabQ project. The Finnish justifications are then mirrored with the findings made on the Swedish interviews (Segerholm, 2009) and other official reports of SNAE.

A methodological opening – ‘What’s the problem represented to be?’ approach combined with the Toulmin model of argumentation

As a methodological opening Carol Bacchi’s (2009) ‘what’s the problem represented to be?’ (WPR) approach is combined with another methodological approach, Stephen E. Toulmin’s (1958) Model of Argumentation. Despite their very different ontological and epistemological origins, the view is taken that both approaches can be used side by side to answer the questions presented above. In addition, the Toulmin model provides a visual and therefore maybe a more approachable way to crystallize the findings of this research.

Drawing upon social constructivism, poststructuralism, feminist theory and Foucauldian governmentality studies, Carol Bacchi’s

(2009) ‘what’s the problem represented to be?’ (WPR) approach suggests that we are governed through problematisations. Rather than investigating problem-solving, Bacchi proposes that we should pay closer attention on problematisation – that is how certain issues become represented as ‘problems’ while other issues do not? These socially constructed ‘problems’ take shape within specific historical, national, international [and local] contexts (Shore & Wright, 1997). Thus, rather than being just exogenous answers, policies (and policymakers) in itself also constitute ‘problems’ – in other words give shape to ‘problems’ and imply by their nature a certain understanding of what needs to change (Bacchi, 2009, p. x–xi).

Stephen E. Toulmin (1922–2009), a British philosopher devoted his works to the analysis of moral reasoning and ethics. In his quest for researching the structure and the logic of argumentation he wanted to extend the analysis to concern practical argumentation, not only theoretical assumptions or examples. Toulmin presented in his most influential work *The Uses of Argument* (Toulmin, 1958) a Model of Argumentation that consists of 6 universal parts for natural claim-making argumentation: claim, data, warrant, backing, rebuttal and qualification (Toulmin, 1958, p. 87–105). Initially intended for the analysis of single sentences, the Toulmin model was only later found useful for representing the structure of macro-level rhetorical and discursive action. The original model has since been revised from various different academic disciplines (see e.g. Hitchcock & Verheij, 2006).

In the studies of political rhetoric and justification by the Toulmin model (e.g. Best, 1987; Kyntäjä, 1993) it has been found useful to operate with the simplified version of three core elements (Figure 1): the *claim* (C) or the conclusion is the outcome of the argumentation that is put forward – in this article the policy practice, whether to publish school specific performance data or not. The claim must be somehow justified – here the *warrants* (W) represent the justifications. For Toulmin a warrant is a value-dependent element of the argument, which works as a bridge connecting the facts or grounds to the claim. Noteworthy is, that many times in a practical argumentation these may remain implicit. The values alone are

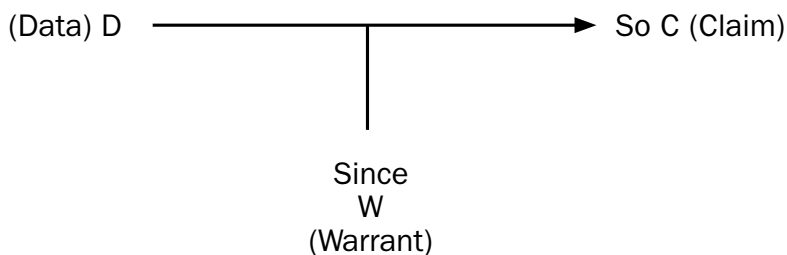


Figure 1. The Toulmin Model of Argumentation (Toulmin, 1958, p. 92)

insufficient, but the claim must be based also on some reasonable grounds, evidence or facts, the *data* (*D*).

However, as Toulmin (1958) notes, the whole argument can be challenged by challenging its elements. Analogically, the elements of the macro-level policymaking reasoning can be critically observed. This is where the Toulmin (1958) model meets Bacchi's (2009) critical WPR approach focusing not only on the *warrants* but also on the *data*. What are the value-dependent justifications behind the opposite policies? On what socially constructed 'facts' are the policies based on – in other words what is the 'problem' represented to be? The Toulmin (1958) model can be seen as an analytical tool in order to clarify how the social reality of the decision-making is constructed – on which "truths", assumptions and values the policymakers build their argumentation and reasoning. In this sense, the analysis is a form of critical discourse analysis, which sees discourses not only as representing the world, but signifying the world, constituting and constructing the world in meaning (Fairclough, 1992).

The results – A combination of different values and 'image of quality'

The WPR approach recommends to conduct the analysis 'backwards' from concrete proposals to reveal what is represented to be the 'problem' within those proposals (Bacchi, 2009, p. 2–3.) In this

chapter the results of the analysis are presented in three consecutive steps (claim, warrants, data) moving the Toulmin (1958) model backwards to answer the two research questions (Figure 2).

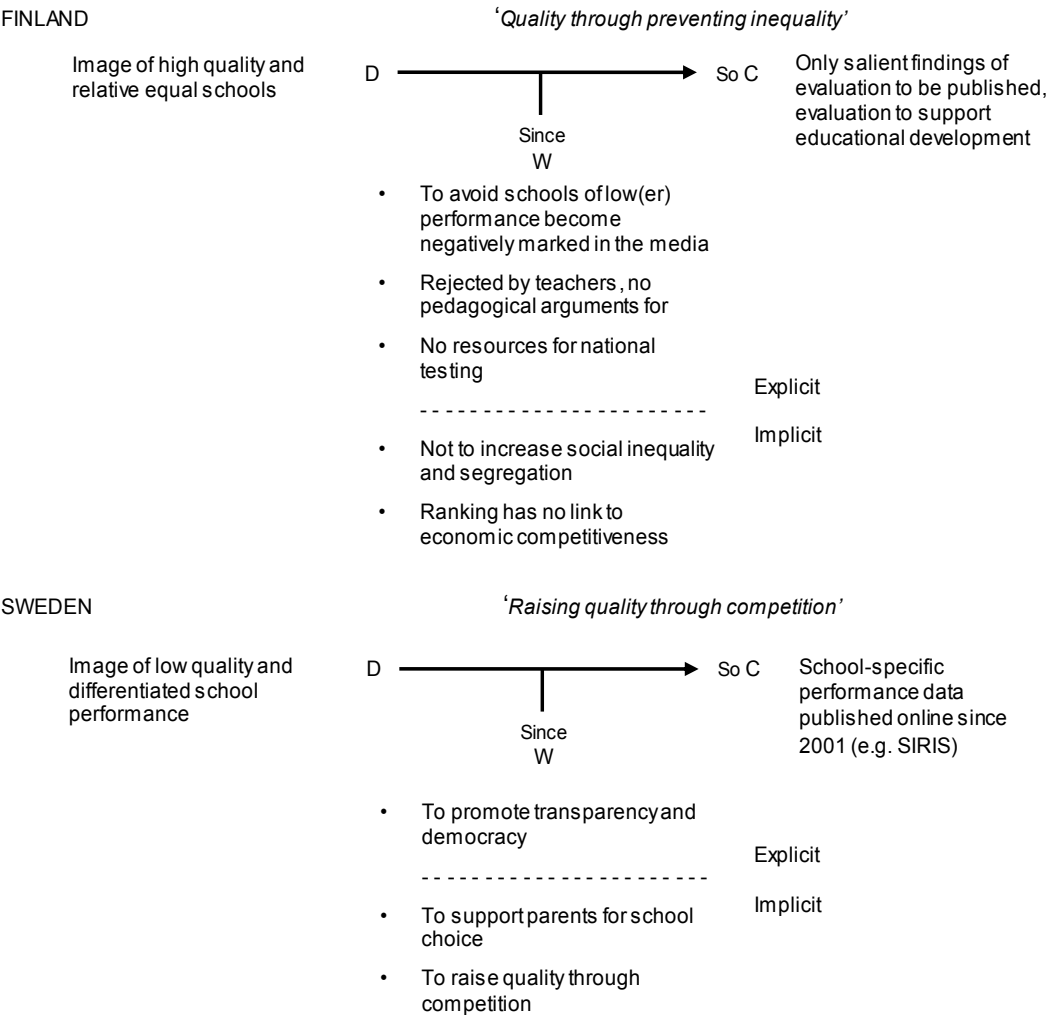


Figure 2. Publication policy of school-specific performance results in Finland and Sweden

The opposite publishing policies – the claim

The Finnish stand not to publish school-specific performance data has been described to be result of both determined policy-making and contingent factors (Simola et al., 2009). The policy actors were well aware of other countries' publication experiences (Jakku-Sihvonen, 2010, p. 321–2; Laukkanen, 1994, p. 103–5), still the decision did not pre-date any wide public discussion that emphasized the role of single policy actors (Syrjänen, 2012). The Finnish stand got its highest legal force and an official form in 1999 in the Basic Education Act:

The purpose of the evaluation of education is to assure that the purpose of this Act is carried out, to support educational development and to improve conditions of learning. ... The salient findings of evaluation shall be published. (Law 1998/628, §21.)

The Act was formulated intentionally implicit leaving the educational decision-makers the possibility to define, what 'the salient findings' to be published, are. However, an explicit position was taken in the report of the Standing Committee for Education and Culture:

The publicity concerns only the main results of evaluations. The purpose of the new Basic Education Act is not to publish information directly linked to an individual school or teacher. Publishing the evaluation results must not in any case lead to the ranking of schools or the categorization of schools, teachers or pupils as weak or good on unfair grounds. (CEC, 1998.)

In Sweden the results of school-specific performance have been published since 2001 on the SNAE webpage through an online system SIRIS (2015). The massive database contains e.g. the municipalities' annual quality reports, leaving certificates from the ninth year of compulsory school, results from nationwide tests and structural facts such as pupil-teacher ratio, teachers' qualifications or costs. Similar online systems providing comparative data on school or municipality level as well as reports with different school/mu-

municipality rankings have since been published also by the Swedish Association of Local Authorities and Regions (2015) and the two teacher unions (Lärarförbundet 2015; Lärarnas riksförbund 2015). SIRIS defines its purpose and target groups as followed:

With SIRIS, we aim to make it easier for schools and municipalities to see what can be improved by examining their own performance and comparing themselves with others ... SIRIS is intended primarily as an aid to everyone who works in schools, to pupils and parents, and to municipal employees and politicians. The Agency wants SIRIS to be a tool for schools to use in their quest for improvement. (SIRIS 2015.)

The Finnish justifications – Fear of media, protecting teachers

To avoid the schools of low(er) performance become negatively branded in the media

The thing that surely caused worry was that it might surface some signs of failure, it might stigmatize some municipalities or schools. But that is why we wrote the law in the way that the schools must not be put in any kind of rank order. That is still in a way distinctive in the Finnish culture not to have these lists...in the tabloids. You have to only go to Stockholm to see every day some list of schools in the newspapers. We took a whole another route here. (FabQ.)

Now the ranking-list was the word we opposed 'til the end, it actually was the specific justification in the decision for conducting sample-based evaluations...Like this one case, [refers to a piece of news on television], showing that this and that one is a poor school, as if it was only information for the parents... (FabQ.)

In the Finnish interviews the most emphatic explicit justifications were in a way or other referred to the role of the media. Both interview citations above show clearly how the suspicion of media misusing school-specific data affected the legislation process. The mistrust fell upon both the print media and the television. The first interviewee expressed how the formulation of the law was a con-

scious choice in order to eliminate the possibility for media to conduct a school ranking. The second interviewee similarly noted that the decision even for choosing the sample-based method instead of national testing was made mostly for the same reason – not to enable ranking lists.

Undoubtedly, the role of the media was seen by the policy actors as powerful but purely negative. As some of the interviewees noted the nature of the sensational reportage on the upper secondary schools' performance in the matriculation exam was not considered as proper journalism and is therefore undesirable to extend any further. The two juridical cases from the beginning of the 2000s were also well in the interviewees' minds.

Rejected by teachers, no pedagogical arguments for

The protective stand was not directed only at schools but to shelter also the teachers from an unwanted stigmatization. It was mentioned e.g. that *'the teachers are frightened and horrified when they look at the school rankings in other countries'* or that *'opposite policy would only upset the Trade Union of Education and the Association of Finnish Local and Regional Authorities, everybody would be just furious...'*

It is important to note that the teachers in the comprehensive school in Finland have enjoyed a relative high status in comparison to many other countries (Simola, 2005b, p. 458–9). In addition, the OAJ is one of the biggest trade unions in Finland and its importance as an influential interest group in the Finnish educational policymaking cannot be neglected – according to Sarjala (2008, p. 145) no educational reforms are made in Finland without the OAJ's approval. In the FabQ project's survey for 1500 teachers the general attitudes towards school ranking were strictly negative – 93 % of the respondent teachers believed that ranking lists would accelerate the polarization between good and poor schools (Rinne et al., 2011, p. 219).

Everyone knows that the time before the matriculation exam is only coaching to pass the final exam, it has not much to do with teaching. And that kind of mentality is just totally unacceptable in the compulsory level. (FabQ.)

Another explicit justification in respect of the teachers was built on the pedagogical basis. The interviewees were very confident that public evaluation results and increasing pupils' testing would affect the pedagogical emphases within the schoolwork negatively. In fear of having poor performance results, the teaching would become narrower and more technical, focusing especially in taking the test. As one of the interviewees mentioned, in the upper secondary school the matriculation exam is a, more or less explicit aim but in the compulsory level the learning and teaching should be something different – more comprehensive and holistic.

No resources for nationwide testing

The question is not how to duplicate 5,000 questionnaire blankets into 50,000, that is possible. But how then to analyze all these blankets properly? There was no other way than this. (FabQ)

...We have estimated the costs many times for national testing and I just can't see how it was supposed to be done ... National testing is so much more expensive. (FabQ)

In Finland the decision to use sample-based evaluation was made during the severe economic recession in the beginning of the 1990s. Many cuts were made in the public sector in Finland, thus at this time it might have been difficult to argue for building up a more resource-demanding agency that would conduct a larger nationwide testing apparatus. However, the justification through lacking resources has a rhetorical nature. The resisting stand continued unchanged despite many years' constant economic growth after the recession.

The implicit warrants

Intertwined with the explicit justifications we may think that at least two implicit warrants backing up the negative stand of the Finnish policy actors. The opposite policy is strongly assumed to increase differentiation between the schools and thus working as a mechanism that increases social inequality and segregation. Moreover, the argument that publishing school performance results and increasing visible competition in education would improve the overall quality of the schools and thereby produce economic competitiveness seems to be widely denied.

The Swedish justifications – Quality through transparency and competition

The interviews with the Finnish and the Swedish policy actors in the FabQ project were not totally uniform which must be taken into account with the interpretation of the results. Nevertheless, the Swedish interviewees emphasized remarkably different issues than the Finnish policy actors. According to Segerholm (2009, p. 205), the Swedish policy actors saw that the recent national QAE activities are well in line with the historical national governance practices as well as with what is promoted internationally. For example, ‘the method of open coordination’ by the European Union as a means to govern was commonly stressed.

In Sweden, the concept of transparency of public institutions has a long history and a status of a self-evident truth. Here, publishing the evaluation results of the compulsory schooling does not make any special exception. The Swedish principle of public access in order to promote transparency and citizens’ democratic rights is stated explicitly also on the SIRIS webpage:

The key social function of schools means that citizens have a democratic right to have access to this information. Child care and education affect almost everyone. In the Agency’s view, public access must therefore be as extensive as possible. (SIRIS, 2015.)

To provide proper information for parental choice was not explicitly mentioned but can be thought as an underlying warrant. Also, as Segerholm notes, the concept of competition is more or less present within the explicit justification of transparency and democratic rights:

Easily understood and commensurable information is paramount to governing. Information is needed for transparency and transparency is needed to supply good examples (best practice) as a basis for improvement, adjustment and competition. Competition is essential in this rationale, locally, nationally and globally. (Segerholm, 2009, p. 205.)

What's the 'problem' – The image of quality

During the 1990s many similar historical events took place in both countries. Many of these interpretations, e.g. decentralization, the economic recession, right-wing governments, membership in the EU or the influence of the OECD etc. may have been adequate to explain nationally the educational reforms, but turn in a comparative research insufficient. How come despite the similarities Finland and Sweden have differentiated in their QAE practices and publication policies? Here, we have to take a look at the contextual frame around the decisions – to find out what is the 'problem' the opposite policies are meant to solve.

In respect of QAE two interrelated distinctions between Finland and Sweden can be found – the image of the quality of the compulsory school system and the logic of improving schools' quality. In Finland the general attitudes towards education and especially to the comprehensive school have been traditionally relative positive. The high quality has been linked to highly qualified teachers, school autonomy, special-needs education and modest testing. In general, there has been a strong conceptual emphasis on the connection of quality and equality (e.g. Rätty, Snellman, Mäntysaari-Hetekorpi, & Vornanen, 1995).

The narrative of the Finnish comprehensive school as a 'success story' (Hämäläinen, Lindström, & Puhakka, 2005) was raised to

another sphere after the release of the OECDs first PISA 2000 results. The continual top positions in the PISA assessment have furthered the global hype around the Finnish compulsory schooling. Most importantly, at the same time the results have provided the Finnish policy actors an undisputed argument to present the national education policy correct and legitimate (Rautalin, 2013) – as if there was no ‘problem’ to be fixed.

In Sweden, just an opposite image started to intensify already during the 1990s, well before the first PISA results (e.g. Björklund et al., 2005, p. 3; Swedish National Agency for Education, 2010). From the middle of the 1990s the new QAE practices, such as the nationwide testing system also made the ‘school failure’ more visible than before (Lundahl, 2002, p. 695). Even though the equality of education was still a priority, the means to ensure both quality and equality was changed to follow the other market-driven reforms and the logic of ‘raising quality through competition’. And if the 1990s was an overture for the declining school performance results, the development during the 2000s has been in Sweden even more alarming. The pupils’ performance results have been declining consistently – which in turn has meant to fix ‘the problem’ by even more testing and standardized evaluation.

In sum, both the Finnish and the Swedish policies can be considered as *consequential to some real issues or experiments* – in other words result of a complex combination of different national governance traditions, different value-related emphases and different contextual policy frames. However, in respect of Bacchi’s (2009) WPR approach, both policies capitalize on a generally and many times unconditionally accepted view that the evaluation results, based either on national tests or international comparisons, do contain evidence-based and objective information on the quality of the education system. If for Sweden the socially constructed ‘problem’ was the decline of the school performance that created the demand for new QAE activities, the Finnish educational policy actors have been *so far* able to ground the QAE policy on the image of high quality in compulsory schooling – even by the cost of limited public transparency.

Discussion

This article has provided a new way to exploit the classical Toulmin (1958) model for conducting comparative and critical policy analysis in the field of education, suitable for other policy fields as well. To end this article a brief hypothetical look in the near future must be taken. As mentioned many times above, the Finnish QAE practices, especially in respect of the publishing policy, have been relatively stable for the last two decades. In this article an explanation has been drawn that on the contrary to Sweden, the practiced policy has been enjoying from the good image of the Finnish compulsory schooling. This good image has been strengthened for the first decade of the 2000s by the Finnish PISA success.

However, the very recent development has left a breach in the image of the Finnish schools' quality. The fresh reports of the PISA 2012 results showed that the Finnish pupils' performance had both fallen significantly from the previous measurements but also the variance between the schools' performance had continued to grow slightly (Kupari et al., 2013). Similar findings have been made also in recent years' national and local evaluations (e.g. Bernelius, 2013; Hautamäki, Kupiainen, Marjanen, Vainikainen, & Hotulainen, 2013). Despite that the Finnish pupils are still performing very well in comparison e.g. to other European countries, the declining performance results – especially in the newsworthy PISA results – may provide a potential turning point in the Finnish education and QAE policy. If the declining trend is to be continued, new openings can be expected even in the consensual and silent policy of not to publish school-specific performance results. However, at the same time the unattractive experiences of the Swedish educational reforms will provide a strong counterargument for the advocates of the present policy.

References

- Allen, R., & Burgess, S. (2011). Can school league tables help parents choose schools? *Fiscal Studies*, 32 (2), 245–261.
- Antikainen, A. (2006). In search of the Nordic model in education. *Scandinavian Journal of Educational Research*, 50 (3), 229–243.
- Bacchi, C. (2009). *Analysing policy: What's the problem represented to be?* Frenchs Forest: Pearson.
- Bernelius, V. (2013). *Eriytyvät kaupunkikoulut – Helsingin peruskoulujen oppilaspohjan erot, perheiden kouluvalinnat ja oppimistuloksiin liittyvät aluevai-
kutukset osana kaupungin eriytymiskehitystä* [Urban schools and segrega-
tion]. Helsinki: Helsingin kaupungin tietokeskuksen tutkimuksia 2013:1.
- Best, J. (1987). Rhetoric in claims-making: Constructing the missing chil-
dren problem. *Social Problems*, 34 (2), 101–121.
- Björklund, A., Clark, M., Edin, P.-A., Fredriksson, P., & Krueger, A. (2005). *The
market comes to education in Sweden: An evaluation of Sweden's surpris-
ing school reforms*. New York: Russell Sage Foundation.
- Bunar, N. (2010). Choosing for quality or inequality – current perspectives
on the implementation of school choice policy in Sweden. *Journal of Edu-
cation Policy*, 25 (1), 1–18.
- CEC. (1998). *Committee for education and culture*. SiVM 3/1998.
- Chubb, J., & Moe, T. (1990). *Politics, markets and America's schools*. Wash-
ington, DC: The Brookings Institution.
- Erkkilä, T. (2010). *Reinventing Nordic openness: Transparency and state in-
formation in Finland* (Doctoral dissertation). Acta politica No. 40. Helsin-
ki: University of Helsinki, Faculty of Social Sciences, Department of Eco-
nomic and Political Studies.
- Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Cam-
bridge: Polity Press.
- Eurydice. (2009). *National testing of pupils in Europe: Objectives, organisa-
tion and use of results*. Brussels: Education, Audiovisual and Culture Ex-
ecutive Agency.
- Fairclough, N. (1992). *Discourse and social change*. Cambridge: Polity Press.
- Friedman, M., & Friedman, R. (1980). *Free to choose*. London: Penguin.
- Hautamäki, J., Kupiainen, S., Marjanen, J., Vainikainen, M.-P., & Hotulainen,
R. (2013). *Oppimaan oppiminen peruskoulun päättövaiheessa: Tilanne
vuonna 2012 ja muutos vuodesta 2001* [Learning to learn at the end of
basic education: Results in 2012 and changes from 2001]. Helsinki:
University of Helsinki, Department of Teacher Education, Research Re-
port No. 347.
- Hitchcock, D., & Verheij, B. (Eds.) (2006). *Arguing on the Toulmin model: New
essays in argument analysis and evaluation*. Dordrecht: Springer.
- Hämäläinen, K., Lindström, A., & Puhakka, J. (Eds.). (2005). *Yhtenäisen pe-
ruskoulun menestystarina* [The success story of the Finnish comprehen-
sive school]. Helsinki: Yliopistopaino.
- Jakku-Sihvonen, R. (2010). Oppimistulosten arviointitiedon puntarointia [On
assessment and evaluation data]. *Hallinnon tutkimus*, 29 (4), 317–324.

- Karsten, S., Visscher, A., & de Jong, T. (2001). Another side to the coin: the unintended effects of the publication of school performance data in England and France. *Comparative Education*, 37 (2), 231–242.
- Kosunen, S. (2014). Reputation and parental logics of action in local school choice space in Finland. *Journal of Education Policy* 29 (4), 443–466.
- Kupari, P., Välijärvi, J., Andersson, L., Arffman, I., Nissinen, K., Puhakka, E., & Vettenranta, J. (2013). *PISA 12 Ensituloksia* [PISA 12 – First results]. Opetus- ja kulttuuriministeriö & Koulutuksen tutkimuslaitos. Helsinki: Opetus- ja kulttuuriministeriön julkaisuja 2013:20.
- Kyntäjä, T. (1993). *Tulopolitiikka Suomessa – Tulopoliittinen diskurssi ja instituutiot 1960-luvulta 1990-luvun kynnykselle* [The Finnish income policy discourse and institutions from 1960s to 1990s]. Helsinki: Gaudeamus.
- Lauder, H., & Hughes, D. (1999). *Trading in futures: Why markets in education don't work*. Buckingham: Open University Press.
- Laukkanen, R. (1994). *Koulutuksen tuloksellisuuden arvioinnin keskeiset kysymykset ja trendit keskushallinnon näkökulmasta* [The main issues and trends in evaluating educational outcomes from the point of view of central administration]. Kasvatustieteiden tutkimuslaitoksen julkaisusarja A. Tutkimuksia 59. Jyväskylä: University of Jyväskylä.
- Law 628/1998. *Perusopetuslaki* [Basic Education Act].
- Linnakylä, P., & Välijärvi, J. (2005). *Arvon mekin ansaitsemme. Kansainvälinen arviointi suomalaisen koulun kehittämiseksi* [International and Finnish evaluation in education]. Jyväskylä: PS-kustannus.
- Lundahl, L. (2002). Sweden: Decentralization, deregulation, quasi-markets – and then what? *Journal of Education Policy*, 17 (6), 687–697.
- Lundahl, L. (2012). A business like any other? The Swedish upper secondary education market in the early 2000s. In J. Kivirauma, A. Jauhiainen, P. Seppänen, & T. Kaunisto (Eds.), *Koulutuksen yhteiskunnallinen ymmärrys – Social perspectives on education* (pp. 89–104). Research in Educational Sciences 59. Jyväskylä: Finnish Educational Research Association.
- Läraryröbundet. (2015). Retrieved from <https://www.lararforbundet.se/>.
- Lärarnas riksförbund. (2015). Retrieved from <http://www.lr.se/>.
- Maw, J. (1999). League tables and the press – value added? *Curriculum Journal* 10 (1), 3–10.
- Nusche, D., Earl, L., Maxwell, W., & Shewbridge, C. (2011). *OECD reviews of evaluation and assessment in education: Norway 2011*. Retrieved from <http://www.oecd.org/edu/school/48632032.pdf>.
- Ozga, J., Dahler-Larsen, P., Segerholm, C., & Simola, H. (Eds.) (2011). *Fabricating quality in education: Data and governance in Europe*. New York: Routledge.
- Power, S., & Frandji, D. (2009). Education markets, the new politics of recognition and the increasing fatalism towards inequality. *Journal of Education Policy*, 25 (3), 385–396.
- Rautalin, M. (2013). *Domestication of international comparisons: The role of the OECD Programme for International Student Assessment (PISA) in Finnish education policy*. Tampere: Tampere University Press.

- Rinne, R. (2000). The globalisation of education: Finnish education on the doorstep of the new EU millennium. *Educational Review*, 52 (2), 131–142.
- Rinne, R., Simola, H., Mäkinen-Streng, M., Silmäri-Salo, S., & Varjo, J. (2011). *Arvioinnin arvo: Suomalaisen perusopetuksen laadunarviointi rehtoreiden ja opettajien kokemana* [Evaluation in the Finnish compulsory schooling: view of the teachers and the principals]. Kasvatusalan tutkimuksia 56. Jyväskylä: Suomen kasvatustieteellinen seura.
- Räty, H., Snellman, L., Mäntysaari-Hetekorpi, H., & Vornanen, A. (1995). Vanhempien tyytyväisyys peruskoulun toimintaan ja koulunuudistuksia koskevat asenteet [Parental satisfaction with the comprehensive school and attitudes towards reforms]. *Kasvatus*, 26 (3), 250–260.
- Sahlberg, P. (2011). *Finnish Lessons: What can the world learn from educational change in Finland?* New York: Teachers College Press.
- Sarjala, J. (2008). *Järki hyvä herätetty: Koulu politiikan pyörteissä* [On school and politics]. Helsinki: Kirjapaja.
- Segerholm, C. (2009). 'We are doing well on QAE': The case of Sweden. *Journal of Education Policy*, 24 (2), 195–209.
- Seppänen, P. (2006). *Kouluvalintapolitiikka perusopetuksessa – Suomalaiskaupunkien koulumarkkinat kansainvälisessä valossa* [School choice policy in comprehensive schooling: School markets in the cities in the international perspective]. Kasvatusalan tutkimuksia 26. Turku: Suomen kasvatustieteellinen seura.
- Shore, C., & Wright, S. (1997). Policy: A new field of anthropology. In C. Shore & S. Wright (Eds.), *Anthropology of policy: Critical perspectives on governance and power*. New York: Routledge.
- Simola, H. (2005a). *Koulukohtaiset oppimistulokset ja julkisuus* [The school-specific performance results and publicity]. *Yhteiskuntapolitiikka*, 70 (2), 179–187.
- Simola, H. (2005b). The Finnish miracle of PISA: Historical and sociological remarks on teaching and teacher education. *Comparative Education*, 41 (4), 455–470.
- Simola, H., Rinne, R., Varjo, J., Pitkänen, H., & Kauko, J. (2009). Quality assurance and evaluation (QAE) in Finnish compulsory schooling: A national model or just unintended effects of radical decentralisation? *Journal of Education Policy*, 24 (2), 163–178.
- SIRIS. (2015). *SIRIS a source of information for understanding of the world of education*. Retrieved from <http://sir.is.skolverket.se/siris/f?p=SIRIS:33:0>
- Swedish Association of Local Authorities and Regions. (2015). Retrieved from <http://skl.se/tjanster/englishpages.411.html>
- Swedish National Agency for Education. (2010). *What influences educational achievement in Swedish schools? A systematic review and summary analysis*. Retrieved from <http://www.skolverket.se/publikationer?id=2318>
- Syrjänen, P. (2012). *Uutta perinnettä rakentamassa – Suomen peruskoulujen otospohjaisen arviointimallin synty politiikan dynamiikan näkökulmasta*

- [The birth of sample-based evaluation in the Finnish compulsory schooling] (Master's thesis). Helsinki: University of Helsinki.
- Telhaug, A.O., Mediås, O.A., & Aasen, P. (2006). The Nordic model in education: Education as part of the political system in the last 50 years. *Scandinavian Journal of Educational Research*, 50 (3), 245–283.
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- van Petegem, P., Vanhoof, J., Daems, F. & Mahieu, P. (2005). Publishing information on individual schools? *Educational Research and Evaluation: An International Journal on Theory and Practice* 11 (1), 45–60.
- Varjo, J. (2007). *Kilpailukykyvaltion koululainsäädännön rakentuminen: Suomen eduskunta ja 1990-luvun koulutuspoliittinen käänne* [Drafting education legislation for the competitive state. The parliament of Finland and the 1990s turn in education policy] (Doctoral dissertation). Research Report 209. Helsinki: University of Helsinki, Department of Education.
- Väljärvi, J. (2012). Suomalainen arviointitodellisuus kansainvälisessä vertailussa [The Finnish evaluation in international perspective]. In J. Kivirauma, A. Jauhiainen, P. Seppänen, & T. Kaunisto (Eds.), *Koulutuksen yhteiskunnallinen ymmärrys – Social perspectives on education* (pp. 171–200). Research in Educational Sciences 59. Jyväskylä: Finnish Educational Research Association
- West, A., & Pennell, H. (2000). Publishing school examination results in England: Incentives and consequences. *Educational Studies*, 26 (4), 423–436.

4. Oppimistulosten kansallisen arvioinnin historiallinen institutionaalistuminen Suomessa ja Ruotsissa

Oppimistulosten mittaamisen käytännöt poikkeavat Suomessa muista Pohjoismaista. Meillä kansalliset arviointikokeet toteutetaan otospohjaisesti eikä arvioinnin tuloksia julkisteta koulukohteisesti. Tässä luvussa vertaillaan perusopetuksen arviointitoiminnan historiallista institutionaalistumista Suomessa ja Ruotsissa. Miten maissa on päädytty päinvastaisiin arviointikäytäntöihin?

Sehän on vieläkin hyvin erikoista, ettei Suomessa ole semmoista kulttuuria ollenkaan, jossa näitä listauksia olisi Iltalehdissä sivukaupalla. Ei tarvitse mennä kuin Tukholmaan, niin näkee joka päivän lehdessä jonkun listan. Se lähti hyvin eri raiteille, mitä tässä.⁹

⁹ Suomalaisen virkamiehen kommentti vuosina 2006–2009 toteutetun tutkimusprojektin *Fabricating Quality in European Education* (Ozga, Dahler-Larsen, Segerholm & Simola 2011) haastatteluaineistossa.

Johdanto

Koulujen oppimistulokset ovat saaneet osakseen kasvanutta huomiota viimeisten 20 vuoden aikana. Euroopassa koulujen julkiarviointi sai alkunsa, kun brittilehdistö ryhtyi vuonna 1992 julkaisemaan kouluviranomaisten tuottamaa arviointitietoa koulujen oppimistuloksista niin sanotuilla ranking-listoilla (*league tables, school performance tables*). Jo vuodesta 1981 lähtien koulujen päättöarvosanojen keskiarvotulokset oli määrätty julkisiksi, mutta vasta 1990-luvun alusta oppimistulokset nostettiin näkyvämmiin kansalaisten tietoisuuteen, kun tuloksista velvoitettiin raportoimaan yhteneväisessä ja vertailtavassa muodossa (West & Pennell 2000). Oppimistulosten julkistaminen Englannissa oli jatkoa oikeistohallitusten 1980-luvulta ajamalle uusliberaalille markkinalogiikan mukaiselle koulupolitiikalle. Koulukohtaisten tietojen julkistamisella oli tässä koulutuksen markkinoitumisessa tietoinen funktio. Sen argumentoitiin tukevan vanhempien vapaampaa kouluvalintaa ja parantavan siten kilpailulogiikan mukaisesti koulujen laatua ja oppimistuloksia kokonaisuudessaan. Koulukohtaisten oppimistulosten julkistaminen suunnattiinkin ensisijaisesti vanhemmille:

Hallituksen näkemyksen mukaan kansalaisten tulee saada kaikki informaatio seuratakseen lapsensa kehitystä, saadakseen tietoa siitä miten koulua johdetaan sekä vertailtavakseen koulua alueen kaikkien muiden koulujen kesken (Department for Education 1994, 3).

Koulukohtaisten tietojen julkistaminen on liitetty tyypillisesti angloamerikkalaiseen koulujen *tilivelvollisuutta* korostavaan arviointipolitiikkaan. Vähemmälle huomiolle on jäänyt, että Pohjoismaista Ruotsissa, Tanskassa ja Islannissa koulujen arviointitiedot on määriteltä julkisiksi¹⁰. Sen sijaan Suomessa koulujen julkiarviointi on kohdannut hyvin voimakasta ja myös konsensuksenomaista vastustusta. Koulujen julkisen vertailun on koettu olevan haitallista, eriyttävän kouluja toisistaan ja heikentävän siten koulutuksen tasa-arvoa.

¹⁰ Myös Norjassa lehdistö on teettänyt koulukohtaisia listauksia julkisuusperiaatteeseen vedoten. Viranomaisraporteissa tulokset ilmoitetaan koulutuksen järjestäjän mukaan, käytännössä kuntakohtaisesti (OECD 2011, 101–2).

Kiistatta suurin vastustus on kohdistunut juuri koulujen välisiin ranking-listoihin, joihin myös valtaosa koulutuspolitiikan tutkimuksista on suhtautunut kritisoivasti. Tässä mielessä muiden Pohjoismaiden harjoittama arviointipolitiikka on herättänyt hämmennystä, etenkin kun juuri koulutuksen tasa-arvon edistäminen on ollut yksi pohjoismaisen peruskoulujärjestelmän keskeisistä tavoitteista (Antikainen 2006; Telhaug, Mediås & Aasen 2006). Toisaalta voidaan ajatella, että arviointitietojen julkistamisen ristiriitaisuus korostuu nimenomaan Pohjoismaissa. Myös viranomaistiedon julkinen saatavuus eli julkisuusperiaate on ollut perustavanlaatuinen osa *pohjoismaisen avoimuuden* hallintokulttuuria (Erkkilä 2010).

Nykymuotoisen arviointitoiminnan linjaukset luotiin 1990-luvun kuluessa. Samassa yhteydessä koulutuksen markkinoituminen, koulujen profiloituminen sekä kouluvalinnan vapauttaminen alkoivat muokata Pohjoismaiden peruskoulujärjestelmiä, näkyvimmin Ruotsissa. Julkisesti rahoitettujen yksityiskoulujen (*friskolor*) lukumäärä ja suosio kasvoivat kouluvalinnan ja voucher-käytännön myötä nopeasti (Björklund, Clark, Edin, Fredriksson & Krueger 2005; Lundahl 2002). Varsin pian myös Ruotsin kansallinen arviointipolitiikka alkoi saada koulujen tilivelvollisuutta korostavia piirteitä. Vuonna 1998 arviointikokeet muutettiin pakollisiksi koko ikäluokalle, ja vuodesta 2001 lähtien koulukohtaiset oppimistulokset on kerätty kaikkien saataville internetiin, kouluviranomaisien ylläpitämiin massiivisiin tietokantoihin. 2000-luvulla arviointikokeiden lukumäärää on entisestään lisätty ja koulutarkastustoimintaa tehostettu. Myös Ruotsin lehdistö on taajaan uutisoinut ja koostanut erilaisia listauksia arvioinnin tuloksista. (Hudson 2007; Rönnerberg, Lindgren & Segerholm 2013; Segerholm 2009.)

Päinvastoin kuin Ruotsissa on Suomessa *kehittävän arvioinnin* tuottama tieto suunnattu pääasiallisesti koulutuksen päätöksentekijöiden käyttöön, eikä yksittäisten koulujen oppimistuloksia ole haluttu julkistaa. Käytännössä koulujen julkinen vertailu on Suomessa mahdotonta, sillä kansallinen oppimistulosten arviointi toteutetaan otospohjaisesti. Toisaalta myöskään suurempaa painetta kiristää kansallista arviointitoimintaa ei Suomessa ole 2000-luvun alussa esiintynyt. Suomalaisoppilaiden PISA-menestys on

vahvistanut harjoitetun koulupolitiikan legitimizeettiä, kun taas Ruotsissa arvioinnin tehostamisella on pyritty korjaamaan heikentyneitä oppimistuloksia ja koulujen eriytymiskehitystä (Wallenius 2015). Tosin aivan viime vuosina samankaltaisia viitteitä oppimistulosten laskusuunnasta on havaittu myös Suomessa, mikä on herättänyt huolta ja keskustelua peruskoulun tulevaisuudesta (ks. esim. Ouakrim-Soivio, Rinkinen & Karjalainen 2015).

Molempien maiden arviointipolitiikka voidaan ymmärtää 1990-luvun rationaalisen päätöksenteon tuotokseksi, joka on rakentunut näkemykselle koulujärjestelmän toimivuudesta. Poliitiikan tutkimuksessa uusinstitutionaalinen lähestymistapa kuitenkin kyseenalaistaa politiikanteon luonteen näin suoraviivaisena ja suunnitelmallisena toimintana. Sen mukaan politiikka ja sitä edustavat yhteiskunnan institutionaaliset toiminnot ovat osin rationaalisen toiminnan, mutta aina myös konteksti- ja kulttuurisidonnaisen toiminnan tuotosta, jotka usein tapahtuvat tiettyjen historiallisten kehityskaarien rajoittamina. Harjoitetulla politiikalla on usein myös odottamattomia seurauksia ja kauaskantoisempia vaikutuksia kuin mihin niillä aikanaan on pyritty. (Pierson 2000.)

Tässä luvussa vertaillaan Suomen ja Ruotsin oppimistulosten arviointikäytäntöjen kehitystä *historiallisen institutionalismin* viitekehyksessä (Mahoney 2000; Pierson 2000; Thelen 1999). Historiallisen institutionalismin mukaan aiemmilla toimilla on taipumus sisältää polkuriippuvuuksia, jotka joko rajoittavat tai vastaavasti mahdollistavat myöhempien kehityskulkujen suuntaa. Tutkimuksessa kysytäänkin, miten maiden oppimistulosten arvioinnin tavoitteet ja käytännöt ennen 1990-lukua näkyvät sen jälkeisessä arviointipolitiikassa. Historiallinen analyysi ulotetaan aina oppimistulosten mittaamisen alkuvaiheisiin, ja se kohdistuu koulutuspoliittisiin dokumentteihin sekä arviointitoiminnan asiantuntijoiden tutkimuksiin. Erityistä huomiota kiinnitetään juuri arvioinnin tiedonkeruumenetelmään sekä tulosten julkistamiseen. Historiallisen analyysin avulla osoitetaan, miten aiemmat toimenpiteet ja päätökset ovat omalta osaltaan mahdollistaneet 1990-luvulla erilaisen suunnan maiden arviointipolitiikoille ja niiden institutionaalisille käytännöille.

Luvun aluksi luodaan katsaus koulukohtaisten tulosten julkistamisen problematiikkaan sekä arvioinnin nykykäytäntöihin Pohjoismaissa, minkä jälkeen siirrytään maakohtaisiin historiallisiin analyysihin Suomessa ja Ruotsissa. Tulosluvun jälkeisessä pohdintaosiossa nostetaan esiin kaksi ajankohtaista koulutuspoliittista keskustelua, joita tarkastellaan kirjoituksessa esitettyjen ajatusten valossa.

Oppimistulosten julkisuus – kolikon kääntöpuolet?

Koulukohtaisten oppimistulosten julkistamisen problematiikkaa on kuvattu saman kolikon kääntöpuoliksi (Karsten, Visscher & de Jong 2001). Tietojen julkistamisen kannattajat ovat perustelleet avointa julkistamispolitiikkaa lähinnä kansalaisten tiedonsaantioikeudella ja hallinnon läpinäkyvyydellä. Juuri hallinnon läpinäkyvyyden edistämisestä on noussut viime vuosikymmeninä eräänlainen yleispätevä tavoite yhteiskuntapolitiikan eri aloilla. Jos sillä aiemmin on viitattu lähinnä kansalaisten demokraattiseen oikeuteen saada tietoa yhteiskunnan keskeisistä toiminnoista, on se uuden julkisjohtamisen (*New Public Management*) myötä kytkeyty entistä vahvemmin ajatukseen taloudellisesta tehokkuudesta ja julkisen sektorin toimintojen tilivelvollisuudesta (Erkkilä 2010).

Tietojen julkistamista on perusteltu kansalaisten oikeudella saada informaatiota verovaroin kustannetun koululaitoksen toiminnasta. Tämän tyyppisen tilivelvollisuusajattelun (*accountability*) korostuminen koulutuksessa on alun perin ollut ominaista juuri angloamerikkalaiselle koulutuspolitiikalle mutta sittemmin levinnyt erilaisin painotuksin myös moniin muihin maihin (Sahlberg 2011). Toinen tyypillinen argumentti koulukohtaisten tietojen julkistamisen puolesta on ollut yhdistää se vanhempien kouluvalintaan. Sekä Suomessa että Ruotsissa vanhempien kouluvalinta-oikeutta laajennettiin 1990-luvulla, joskin valinnan merkitys on Ruotsissa saanut Suomeen verrattuna isomman roolin niin sanottujen vapaakoulujen (*friskola*) perustamisen myötä (Seppänen 2006).

Yleisesti käytetty argumentti onkin, että vanhemmilla on oikeus saada relevanttia tietoa juuri eri koulujen oppimistuloksista valintansa tueksi. Vertailtavassa muodossa esitetyistä oppimistuloksista on näin muodostunut eräänlainen mittari, jonka mielletään usein kuvaavan yksittäisten koulujen laatua.

Kolikon kääntöpuolena on koulukohtaisten tietojen julkistamista kritisoitu merkittävistä haittavaikutuksista, ja varsinkin koulujen ranking-listaus on kohdannut voimakasta kritiikkiä eri maissa. Etenkin monet koulutuksen tutkijat sekä myös koulujen opettajat ovat vastustaneet avointa julkistamispolitiikkaa. Adrie Visscherin (2001) mukaan kritiikki on ollut kolmenlaista; se on kohdistunut 1) *teknisanalyyttisiin*, 2) *tiedon käytettävyyden* sekä 3) *yhteiskunnallis-eettisiin* ongelmiin. Näistä ensin mainitulla tarkoitetaan muun muassa julkistettavien tietojen validiuteen ja ranking-listojen yksinkertaistavaan luonteeseen liittyviä ongelmia. Samojen koulujen tulokset ovat saattaneet vaihdella huomattavasti vuodesta toiseen, samoin koulujen keskinäiset järjestykset ovat olleet täysin riippuvaisia siitä, mitä asioita vertailuun on milloinkin sisällytetty. Ennen kaikkea tulosten on nähty kertovan pikemmin koulujen oppilasaineksesta kuin itse opetuksen laadusta. Tähän kritiikkiin on pyritty vastaamaan niin sanottujen *value added* -mittareiden kehittelyllä, jotka pyrkivät huomioimaan tuloksissaan esimerkiksi koulun oppilaspohjan, alueen tai perheen sosioekonomisen taustan ja oppilaiden lähtötason, jolloin tulosten voidaan ajatella kertovan koulun vaikutuksesta oppilaiden suoriutuksiin. (Simola 2005; Visscher 2001.)

Mittareiden tekninen kehittely ei kuitenkaan ole ratkaissut kysymystä tietojen *käytettävyydestä* tai tulosten julkistamisen *eettisistä* seuraamuksista. Esimerkiksi tietojen tasa-arvoinen käytettävyys on kyseenalaista, kun kaikilla perheillä ei ole samanlaisia resursseja tulkita, mitä eri tiedot sisältävät tai mistä niiden oletetaan kertovan. Kaikkein kovaäänisin kritiikki on kohdistunut kuitenkin *yhteiskunnallis-eettisiin* epäkohtiin. Koulutyön kannalta tulosten julkistamisen on arvioitu kaventavan opetustyötä yksipuolisesti arviointikokeissa suoriutumiseen (*teaching to the test*) sekä aiheuttavan tarpeetonta stressiä niin oppilaille kuin opettajille. Sen on myös nähty heikentävän koulutuksen toimijoiden välistä

luottamusta. Ennen kaikkea, kilpailulogiikan mukainen oletus koulutuksen laadun parantumisesta ei ole vakuuttanut moniakaan koulutuspolitiikan tutkijoita. Päinvastoin, tietojen julkistamisen on koettu pikemmin kiihdyttävän koulujen välisten tasoerojen kierrettä ja vaikuttavan kaikkein haitallisimmin juuri haasteellisimmissa toimintaolosuhteissa toimiviin kouluihin (esim. alemman sosioekonominen alueen koulut tai koulut, joissa on suuri maahanmuuttajien osuus). Käsité *naming and shaming* kuvastaa juuri tämän kaltaista kierrettä, jossa sekä oppilaat että opettajat alkavat vältellä julkisuudessa leimattuja huonomaineisia kouluja. (Karsten ym. 2001; Maw 1999; van Petegem, Vanhoof, Daems & Mahieu 2005; Power & Frandji 2009; Simola 2005; West & Pennell 2000.)

Asia, jota monissakaan ulkomaisissa tutkimuksissa ei ole erikseen huomioitu, koskee kansallisten oppimistulosten tiedonkeruumenetelmää. Useimmissa Euroopan maissa oppimistulosten mitaamiseen osallistuvat kaikki ikäluokan oppilaat (Eurydice 2009a). Kenties tästä syystä koko ikäluokan testaamista on käsitelty itsensänselvyytenä ja huomio on kiinnittynyt tulosten julkistamisen hyöty- ja haittavaikutusten arviointiin. Yksinkertaisin ratkaisu tietojen julkistamisen dilemmaan – se, ettei julkistettavaa dataa ole saatavilla ensinkään – on usein sivuutettu. Kun arviointikokeisiin osallistuu vain murto-osa maan kouluista ja oppilaista ja otokseen valitut koulut vaihtelevat arviointikokeittain, ei tulosten julkaisu ole kovin mielekäästä – jos siis laisinkaan. Julkistamispolitiikassa on siten lopulta aina kyse siitä, mitkä yhteiskunnan tahot määritellään arviointitiedon käyttäjiksi ja mitä arviointitiedolla tavoitellaan.

Oppimistulosten kansallinen arviointi ja tulosten julkisuus Pohjoismaissa – Suomi poikkeustapaus?

Tässä luvussa luodaan katsaus kansallisiin oppimistulosten arviointikäytäntöihin Pohjoismaissa. Kansallisilla arviointikokeilla (*national testing, standardized testing, national assessment*) tarkoitetaan

tässä kokeita, joihin osallistuvat oppilaat tekevät saman kokeen, jonka tulokset ovat keskenään vertailukelpoisia¹¹. Useimmissa maissa säännöllistä arviointia toteutetaan äidinkielessä, matematiikassa ja englannin kielessä, muissa oppiaineissa usein harvemmin. Tässä yhteydessä on syytä muistaa, että oppimistulosten kansallinen arviointi on vain osa arviointitoiminnan laajaa kenttää (esimerkiksi koulujen suorittama itsearviointi, kunnan teettämä paikallinen arviointi, mahdollinen koulutarkastustoiminta, kansainväliset arvioinnit).

Taulukon 1 tiedot perustuvat Eurydice-tiedonvaihtoverkoston¹² teettämään laajaan tutkimukseen kansallisista oppimistulosten arviointikäytännöistä 30 Euroopan maassa (Eurydice 2009a). Tutkimusraportin mukaan eurooppalaiset arviointikäytännöt ovat hyvin vaihtelevia, ja maakohtaisia eroavaisuuksia esiintyy niin arviointimethodissa, arvioinnin tarkoituksessa sekä ylipäänsä arviointitiedolle annetussa painoarvossa. Erilaisista arviointikäytännöistä huolimatta yhteistä kaikille maille on, että oppimistulosten seuranta – tavalla tai toisella – muodostaa tärkeän informaatio-ohjauksen välineen, joka liittyy läheisesti opettamiseen, oppimiseen ja siten myös koulutuksen laadunarviointiin. (Eurydice 2009a, 11, 63–66.) Raportin pohjalta koostettuun taulukkoon on kerätty tutkimuksen kannalta olennaiset vertailutiedot: 1) minkälaisia pääasiallisia tavoitteita kansallisen tason oppimistulosten arvioinnille on asetettu, 2) miten arviointitietoa kerätään ja 3) miten arviointitietoa julkistetaan.

Maakohtaisista eroavaisuuksista huolimatta näyttäisi suomalaisen arviointitoiminta poikkeavan muista Pohjoismaista kaikkein selvimmin. Kun Suomessa kansallisten arviointikokeiden tehtävänä on tuottaa informaatiota lähinnä koulujärjestelmätasolla, on

¹¹ Tanskassa käytetty tietokoneavusteinen ns. adaptiivinen oppimistesti muuntaa tehtäväsarjan vaikeusastetta oppilaan vastausten perusteella eikä siten ole kaikille oppilaille täysin yhteneväinen (Eurydice 2009a, 36).

¹² Vuonna 1980 perustettuun Eurydice-verkoston kuuluivat vuonna 2013 EU:n 28 jäsenvaltion lisäksi Norja, Islanti, Liechtenstein, Sveitsi, Makedonia, Montenegro, Serbia ja Turkki. Euroopan komission hallinnoiman verkoston päätehtävänä on tuottaa vertailukelpoista tietoa eri maiden koulutusjärjestelmistä ja koulutuspolitiikasta. (http://eacea.ec.europa.eu/education/eurydice/index_en.php)

Taulukko 1. Perusopetuksen oppimistulosten kansallinen arviointi Pohjoismaissa (Eurydice 2009a, 25, 27, 56, 76, 98–99, 103–104; OECD 2011, 101–102)

		Islanti	Norja	Tanska	Ruotsi	Suomi
Arviointikokeiden pääasialliset ta- voitteet	Oppilasarviointia/jatko- koulutusta koskeva päätöksenteko		•	•	•	
	Koulujen/koulutus- järjestelmän seuranta	•	•		•	•
	Oppimisvaikeuksien havaitseminen	•	•	•	•	
Tiedonkeruu- menetelmä	Arviointikokeet pakollisia kaikille kouluille ja oppilaille	•	•	•	•	
	Otospohjaiset arviointikokeet					•
Tulosten julkistaminen	Koulukohtaiset oppimis- tulokset julkisia	•	•	•	•	

kokeilla muissa maissa myös oppilaita ja opettajia koskevia käyttö-tarkoituksia. Koulujärjestelmän seurannan ohella kokeiden tulokset voivat vaikuttaa opettajien antamiin arvosanoihin ja siten oppilaiden koulu-uraan. Lisäksi kaikissa muissa Pohjoismaissa kuin Suomessa pyritään kokeilla havaitsemaan mahdollisia oppimisvaikeuksia. Esimerkiksi Tanskassa opetuksen kuluessa toteutettava formatiivinen arviointi (Jakku-Sihvonen 2010, 318) seuraa hyvin läheltä yksittäisen oppilaan oppimisen edistymistä. Tanskassa arvioinnin painopiste onkin koulujärjestelmän sijasta oppilaskohtaisessa arvioinnissa, jossa hyödynnetään paljon sähköisiä koepankkeja sekä tietokoneella suoritettavia testejä¹³.

Arvioinnin tavoitteet konkretisoituvat tiedonkeruumenetelmässä ja tulosten julkisuudessa. Ainoana Pohjoismaana Suomi toteuttaa oppimistulosten kansallisen arvioinnin otospohjaisesti. Kansallisiin arviointikokeisiin poimittava alueellisesti edustava otos kattaa noin 5–10 prosenttia mitattavasta ikäluokasta ja noin 15 prosent-

¹³ Arviointikokeiden tehtävät vaihtelevat myös oppilaiden iän mukaan. Pääsääntönä voi sanoa, että alaluokilla korostuu diagnostinen oppimisvaikeuksien havainnointi, kun taas peruskoulun päättövaiheessa kokeet vaikuttavat oppilaiden arvosanoihin.

tia maan kouluista eli noin 4 000–6 000 oppilasta (Ouakrim-Soivio 2013, 20). Kaikissa muissa Pohjoismaissa arviointeihin osallistuu koko ikäluokka. Islannissa, Tanskassa ja Ruotsissa koulujen arviointitulokset ovat lähtökohtaisesti julkisia. Eurydice-raportin ilmestymisvuoden jälkeen myös Norjassa arviointituloksia on luovutettu medialle, tosin vasta erikseen pyydettyäessä (OECD 2011, 101–2). Näyttäisi siis siltä, että ainakin Pohjoismaissa koko ikäluokan kattava kansallinen arviointi yhdistyy oppimistulosten julkisuusperiaatteeseen.

Tutkimuskysymys – miten päinvastaisiin arviointikäytäntöihin päädyttiin?

Arvioinnin painoarvo koulutuksessa kasvoi Suomessa ja Ruotsissa, kun päätöksentekovaltaa siirrettiin 1980- ja 1990-lukujen taitteessa paikallistasolle. Oppimistulosten kansallisen arvioinnin oli määrä varmistaa, että koulutukselle asetetut tavoitteet toteutuisivat eri kunnissa. Kuten edellä esitettiin, päätyivät maat harjoittamaan silmiinpistävän erilaista arviointipolitiikkaa. Suomessa Opetushallituksen pitkäaikaisella toimijalla Ritva Jakku-Sihvosella oli suomalaisen linjauksen valmistelussa keskeinen rooli:

Olin asian vastuullinen valmistelija, ja päädyin esittämään otanta-perustaista, arviointikohteiden vuosittaiseen vuorotteluun perustuvaa järjestelmää siksi, että se täyttää hallinnon edellyttämät taloudellisuuden, tehokkuuden ja vaikuttavuuden vaatimukset sekä estää oppimistulosten arviointien pohjalta laadittavat ranking-listat, joiden kielteisistä vaikutuksista tein havaintoja ulkomailla käytössä oleviin arviointijärjestelmiin tutustuessani. Opetushallituksen johto vahvisti esitetyt periaatteet perusopetuksen kansallisen arviointijärjestelmän kehittämisen lähtökohdiksi, ja ensimmäiset järjestelmän mukaiset arvioinnit toteutettiin vuonna 1998. (Jakku-Sihvonnen 2013, 20.)

Ruotsissa määrättiin kansalliset arviointikokeet pakollisiksi koko ikäluokalle vuonna 1998. Vain muutamaa vuotta myöhemmin Ruotsin koulutuksen kansallinen keskusvirasto *Skolverket* kokosi

koulukohtaiset arviointitulokset ylläpitämiinsä internet-tietokantoihin, SIRISIin ja SALSAan, kaikkien vapaasti saataville. Tulosten julkistamiskäytännön perusteluissa on havaittavissa Suomeen verrattuna täysin vastakkainen lähestymistapa. Julkisuusperiaatteen ja kansalaisten demokraattiseen oikeuteen vedoten arviointitietojen tulee olla mahdollisimman kattavasti saatavilla:

Tavoitteiden saavuttaminen edellyttää tietoa nykytasosta ja siitä, mitä kuuluu parantaa. Skolverket loi SIRISin edistääkseen tämän informaation saatavuutta koulutuksen ja päivähoidon osalta. Koska koululla on keskeinen funktio yhteiskunnassa, on sitä koskevan tiedon saatavuus kansalaisten demokraattinen oikeus. Lasten päivähoito ja koulutus koskettaa lähes jokaista. Täten, tiedon julkisuuden tulee viraston näkemyksen mukaan olla mahdollisimman kattavaa, koskien etenkin tietoa eri koulujen välillä ilmenevistä laatueroista. (Skolverket.)

Suomalaisesta perspektiivistä katsottuna koulukohtaisten tietojen julkaisu Ruotsissa saattaa herättää hämmennystä, etenkin kun sen moninaiset koulujen eriarvoistumista kasvattavat haittavaikutukset ovat meillä yleisesti tunnustettuja. Perustelun yhtymäkohdat uusliberaaliin koulujen välistä kilpailua, avointa vertailua ja tilivelvollisuutta korostavaan politiikkaan ovat koulujen laatueroista puhuttaessa varsin ilmeiset.

Miten vastakkaiset arviointikäytännöt kahden Pohjoismaan välillä ovat selitettävissä? Erilaisia tulkintoja voitaisiin antaa monista eri näkökulmista. Tässä luvussa ollaan kiinnostuneita oppimistulosten arvioinnin historiallisista kehityskaarista ja ajasta, jolloin koulukohtaisten tulosten julkisuus tai ranking-listaukset eivät vielä olleet koulutuspoliittisen keskustelun keskiössä. Tämä edellyttää arviointitoiminnan historiallisen institutionaalistumisen tuntemista.

Historiallinen institutionalismi teoreettisena viitekehyksenä

Politiikan tutkimuksessa 1970-luvun loppupuolella alkanut uusinstitutionaalinen lähestymistapa kyseenalaistaa politiikanteon luonteen puhtaasti funktionaalisena, laskelmoivana ja rationaalisena prosessina. Se ei kiistä, etteivätkö poliittiset päätökset, esimerkiksi Suomessa linjaus otospohjaisesta arvioinnista, olisi tietoisesti harkittuja toimia. Uusinstitutionalistien mukaan politiikkaa tehdään ja tulee siten myös tutkia sen yhteiskunnallisessa, kulttuurisessa ja historiallisessa kontekstissa. Päätöksillä on monesti myös täysin odottamattomia ja huomattavasti kauaskantoisempia seurauksia kuin mitä päätöksentekohetkellä on välttämättä osattu kuvitella. Uusinstitutionaalinen lähestymistapa sisältää sosiologista, historiallista, rationaalista ja diskursiivista lähestymistapaa korostavia koulukuntia, mutta painotuserot näiden kesken ovat häilyviä ja käytännön tutkimuksessa usein yhdistellään eri koulukuntien selitysmalleja (Erkkilä 2010; Pierson 2000; Thelen 1999).

Historiallisen institutionalismin viitekehyksessä niin sanottu *polkuriippuvuuden* käsite on keskeinen. Polkuriippuvuudella tarkoitetaan ajatusta siitä, että aiemmin tapahtuneilla toimenpiteillä ja päätöksillä on tapana kertautua, tuottaa tietynlaista toimintaa ja vaikuttaa tavalla tai toisella myöhempien kehityskulkujen todennäköiseen suuntaan. Mitä vakiintuneemmat ja pidemmät historialliset institutionaaliset funktiot tietty yhteiskunnan toiminto omaa, sitä vaikeampaa sen muuttaminen nopeasti tai radikaalisti on. Polkuriippuvuutta ei silti tule ymmärtää deterministisenä, toimintaa tiettyyn suuntaan ennalta määräävänä tekijänä. Pikemminkin sen valossa voidaan esittää hypoteeseja siitä, olisiko jokin erilainen kehityskulku ollut todennäköinen ilman tiettyä historiallista kehitystä (Pierson 2000).

Sattuman merkitystä korostavat *kontingentit* tekijät tai tapahtumat voivat muuttaa kehityskulkujen suuntaa. James Mahoney (2000, 507) mukaan juuri erilaiset kontingentit tekijät saavat liiketä institutionaalistumisprosessiin. Koulujen ranking-kulttuurin synty Englannissa 1990-luvun alussa sekä myöhemmin 2000-luvulla

molempien maiden PISA-tutkimusten tulokset ymmärretään arviointipolitiikan keskeisiksi taitekohdiksi (*critical juncture*), joissa kontingentit tekijät voivat saada aikaan muutosta. Voidaan ajatella, että uusi julkiarvioinnin käytäntö Englannissa ”pakotti” myös muut maat arvioimaan uudessa valossa omaa arviointipolitiikkaansa ja julkisuuskysymystä. 2000-luvulta lähtien puolestaan PISA-tutkimus on muokannut vahvasti koulujärjestelmän ja peruskoulupolitiikan kontekstia. Suomessa jossain määrin yllätykselliset kärkisijat on artikuloitu suomalaisen koulupolitiikan menestystarinaksi, kun taas Ruotsissa vähintään yhtä yllättävä huomio maan oppimistulosten heikosta tasosta kasvatti painetta arviointitoiminnan tehostamiselle monin eri tavoin. Samat kontingentit tekijät eivät kuitenkaan välttämättä johda eri historiallisissa ja kulttuurisissa konteksteissa samanlaiseen lopputulokseen (Pierson 2000). Oppimistulosten heikkenemisen ei siis pidä ajatella automaattisesti johtavan kaikkialla arviointitoimien lisäämiseen, mutta se saattaa lisätä niiden todennäköisyyttä.

Luvussa siirrytään seuraavaksi tarkastelemaan, minkälaisten tavoitteiden ympärille oppimistulosten kansallinen arviointi Suomessa ja Ruotsissa syntyi ja miten institutionaalisiksi toimintatavoiksi vakiintuneet arviointikäytännöt ovat omalta osaltaan mahdollistaneet nykymuotoisten ja maille vastakkaisten arviointilinjausten toteutumisen.

Oppimistulosten kansallisen arvioinnin institutionaalistumisen kehitys Ruotsissa 1930–1990

Oppilaiden oppimistuloksien mittaamisen syntyhistoria ajoittuu 1900-luvun alkupuolelle. Ensimmäisen maailmansodan aikaan Yhdysvalloissa maan sotilasjohto oli alkanut hyödyntää erilaisia psykometriikkaan ja kvantitatiivisiin tutkimusmenetelmiin nojavia älykkyys- ja soveltuvuustestejä rekrytoinnissaan. Sodan päätyttyä vastaavanlaiset menetelmät levisivät koulumaailmaan, kun

oppilaiden oppimisen kehitystä ryhdyttiin mittaamaan standardisoiduilla oppimistesteillä. (Shiel, Kellaghan & Moran 2010, 38–9.)

Jos Yhdysvaltoja voidaan pitää oppimistulosten mittaamisen syntymaana, on Ruotsi yksi arviointitoiminnan edelläkävijämaista Euroopassa. Kansallisen arvioinnin juuret ulottuvat Ruotsissa aina 1930-luvulle eli huomattavasti pidemmälle kuin Suomessa. Uudenlaisten arviointikokeiden tarve ilmeni ennen kaikkea koulun sisällä. Oppilaan jatko kansakoulusta (*folkskola*) oppikouluun (*realskola*) oli perustunut siihen asti äidinkielessä ja laskennossa suoritettavan sisäänpääsykokeen tulokseen. Valintakokeen koettiin sisältävän kuitenkin monia epäkohtia. Monet opettajat olivat kritisoineet, kuinka heidän käsityksenä mukaan osa kokeessa hylätyistä oppilaista olisi tiedoiltaan ja taidoiltaan ollut oikeutettuja jatkokoulutukseen – vastaavasti yksittäisessä kokeessa oli saattanut onnistua osaamistasoltaan heikompi oppilas. Kokeen nähtiin lisäksi aiheuttavan paineita niin oppilaille kuin heidän perheilleen ja lisäävän epätasa-arvoa, kun osassa maata oli alettu järjestää koulutyön ohella ylimääräisiä kokeeseen valmistavia valmennuskursseja. Yksittäiseen kokeeseen perustuvaa pääsykoejärjestelmää pidettiin ylipäänsä epäoikeudenmukaisena ja lapsen ikävaiheeseen sopimattomana valintamenetelmänä. (Fredriksson 1950, 171–2; Ljung 2000, 8.)

Kritiikki johti lopulta vuonna 1938 kouluhallituksen raporttiin, jossa todettiin, että pääsykoejärjestelmästä tulisi luopua ja oppikouluun pääsyn tulisi perustua oppilaan kouluarvosanoihin. Jotta opettajien oppilaille antamat arvosanat olisivat koko maassa mahdollisimman yhteismitallisia, tulisi arvosanojen antamisessa hyödyntää yhteisiä standardisoituja kokeita (SOU 1938:29). Standardisoitujen arviointikokeiden (*standardprov*) synty Ruotsissa oli siis suoraan sidoksissa oppilaiden opintomenestyksen arvioimiseen ja jatkokoulutukseen hakeutumiseen liitettyyn problematiikkaan, ja yhteisten kokeiden oli määrä kaventaa oppilasarvioinnissa havaittua opettajien ja koulujen välistä virhemarginaalia. Matematiikan arviointikokeiden parissa mittavan työuran tehneen pedagogiikan professori Bengt-Olov Ljungin (2000) mukaan oppimistulosten arvioinnin lähtökohtainen tavoite – miten ratkaista oppilasarvioinnin

oikeudenmukainen toteutuminen kaikkien oppilaiden kesken – on pysynyt Ruotsissa periaatteessa muuttumattomana kohta jo lähes sadan vuoden ajan (Ljung 2000, 7).

Ensimmäinen kansallinen standardikoe otettiin käyttöön usean vuoden kehitystyön päätteeksi lukuvuoden 1943–44 keväällä. Kokeita alettiin järjestää aluksi kansakoulun toisella, neljännellä ja kuudennella luokalla äidinkielessä ja laskennossa, mutta muutama vuoden kuluttua toisen luokan testeistä luovuttiin. Lukuvuodesta 1949–50 lähtien oppilaiden jatko oppikouluun määräytyi annettujen todistusarvosanojen perusteella (Ljung 2000, 10–13).

Tutkimuksen kysymyksenasettelun kannalta tulee tehdä kaksi olennaista huomiota. Ensinnäkin, standardikokeen tuloksia ei arviointitoiminnan alkuvaiheessa millään muotoa tarkoitettu julkisiksi. Koetulokset oli tarkoitettu ensisijaisesti opettajan käyttöön. Toiseksi, standardikokeiden käyttö ei ollut pakollista, vaan pikemmin kouluviranomaisten suositus opettajille:

Asiantuntijoiden mukaan kokeen ei tule olla pakollinen, vaan jokaisen opettajan tulee yhdessä koulun johdon kanssa päättää, miten koetta haluaa käyttää (SOU 1945: 45, 66).

Vapaaehtoisuudesta huolimatta – tai ehkä juuri siksi – opettajakunta Ruotsissa otti tarjotun uudistuksen hyvin innokkaasti vastaan. Opettajien suhtautumista standardikokeeseen voisi kuvailla varsin pragmaattiseksi, sillä se miellettiin ennen kaikkea oikeudenmukaisen arvioinnin apuvälineeksi. Ljungin (2000, 7) mukaan opettajien ohella niin oppilaat kuin heidän vanhempansakin vaativat oikeudenmukaisuuden toteutumista arvioinnissa. Standardikokeella oli siten alusta asti laajalti hyväksytty ja legitiimi asema ruotsalaisessa koulujärjestelmässä.

Kansa- ja oppikoulun korvaaminen peruskoulujärjestelmällä vuonna 1962 (siirtymäaika vuoteen 1972) ei muuttanut standardikokeen tarkoitusta. 1960–70-luvuilla tehtiin kokeeseen muutamia tarkennuksia esimerkiksi vaihtamalla oppiaineiden arvioinnin ajankohtia eri ikäluokkien kesken. Peruskoulu-uudistus laajensi standardikokeen vuonna 1962 ensimmäistä kertaa tarjolle myös yläasteen opettajille. Koe järjestettiin äidinkielessä, englannissa ja

matematiikassa. Opettajien jatkuneesta myönteisestä suhtautumisesta kertoo se, että 86 prosenttia eli noin 30 000 oppilasta osallistui kokeisiin (Ljung 2000, 16). Vähitellen kokeen käyttö painottui yläluokille, etenkin kun vuosien 1969 ja 1982 välisenä aikana Ruotsissa luovuttiin arvosanojen antamisesta ala-asteella:

Muutokset oppilaiden arvioinnissa vähensivät opettajien kiinnostusta standardikokeen käyttöön, joka siihen mennessä oli kokeen vapaaehtoisuudesta huolimatta ollut lähes sataprosenttista. Arvosanojen käytöstä ala-asteella luovuttiin kokonaisuudessaan kuten myös yksittäisissä kunnissa keskiluokilla (*mellanstadiet*). Myös yläasteella kokeen käyttö väheni, joskin pienemmässä määrin. (Henricson 1987, 9.)

Tiivistäen voidaan todeta, että oppimistulosten kansallisen arvioinnin peruseräpäätteet, kokeiden käytön vapaaehtoisuus ja tulosten hyödynnettävyys ensisijaisesti opetustyössä, pysyivät lähes 50 vuoden ajan ennallaan. Standardikokeiden järjestämisen yhteydessä opettajia toistuvasti ohjeistettiin, että kokeen tarkoituksena oli antaa opettajille käsitys oman luokkansa osaamistasosta eikä se saisi lisätä kokeeseen liittyvää jännitystä tai vaikuttaa opetuksen sisältöön niin, että testattavaa ainetta erityisesti harjoiteltaisiin. Juuri tästä syystä koetulosten luottamuksellinen käyttö nähtiin hyvin tärkeäksi. (Ljung 2000, 16–17.) Arviointikokeiden tulosten laajempaan julkistamiseen ei siis vielä tässä vaiheessa kohdistunut ulkopuolisia vaatimuksia (Henrysson 1969).

Oppimistulosten mittaamisen juuret Suomessa peruskoulu-uudistuksessa

Suomessa oppimistulosten kansallinen arviointi käynnistyi huomattavasti Ruotsia myöhemmin ja on vasta viime vuosikymmeninä saanut systemaattisemman muodon. Ruotsin tavoin psykometrisen ja positivistisen tiedekäsityksen esiinnousua ilmensivät muun muassa jo 1900-luvun alussa koulumaailmaan soveltuvien älykkyystestien kehittäminen (Rinne, Kivirauma & Lehtinen 2004) sekä ylipäänsä psykologian synty akateemisena oppiaineena

1930-luvulla. Hallinnollisesti koulun toiminnan arviointi nojasi kuitenkin 1900-luvun rinnakkaiskoulujärjestelmässä pitkälti alueelliseen koulutarkastustoimintaan. Vasta peruskoulu-uudistuksen yhteydessä 1960- ja 1970-lukujen taitteessa alkoi tutkimustietoon perustuva oppimistulosten mittaaminen vähitellen saada institutionaalisia piirteitä koulutuksen hallinnallisessa ohjauksessa (Laukkanen 1998; Lyytinen & Lukkarinen 2010; Välijärvi 2008).

Suomi oli 1960-luvulla osallistunut kansainvälisiin oppimistuloksia ja -asenteita mittaaviin tutkimuksiin, ja kansallinen oppimistulosten arviointitoiminta kehittyikin osittain näiden kokemusten pohjalta (Leimu 2004). Jyväskylän yliopiston yhteyteen perustettu Kasvatustieteiden tutkimuslaitos alkoi Kouluhallituksen toimeksiannosta tuottaa niin sanottuja koulusaavutuskokeita, joista käytettiin yleisesti nimitystä *yhteiset kokeet*. Ensimmäiset kokeet järjestettiin vuonna 1967 kokeiluperuskoulun yläasteella vieraissa kielissä, sitä seuraavana vuonna äidinkielessä ja matematiikassa, ja sittemmin myös muissa aineissa (Linnakylä 1974). Hannu Saaren (1983) mukaan kerätyn arviointitiedon käyttötarkoitus oli tarkasti rajattu ja sen oli määrä antaa 1) opettajille käsitys oman luokkan-
sa osaamistasosta suhteessa valtakunnalliseen tai alueelliseen tasoon, 2) viranomaisille yleisluontoista tietoa koululaitoksen tuotoksista sekä opetussuunnitelman kehittämistarpeista ja 3) palvel-
la tiedeyhteisöä sitomalla koulusaavutuskokeet yleisemmin valtakunnalliseen koulutuksen tutkimukseen.

Kokeiden tarkka kohdennettu käyttö sisälsi Saaren mukaan jo tuolloin näkemyksen siitä, ettei koulusaavutuskokeiden ole määrä johtaa kouluja tai opettajia kontrolloivien toimien kasvuun tai valtakunnallisten päättökokeiden kehittämiseen:

Mihin koulukokeita ei pidä käyttää. Yleisesti ottaen koulukokeita ei pidä käyttää mihinkään muuhun. ”Muu” sisältää esim. opettajan ja/tai koulun tuotosten hallinnollisen kontrollin, kokeille ei pidä antaa valtakunnallisen tutkinnon leimaa eikä tehtävää – sen vuoksi kokeita ei voi käyttää myöskään valinta- tai pääsykoetyyppisesti, kokeita ei pitäisi käyttää kaupallisesti, liiketoiminnallisin periaattein. Koulukokeista ei pidä tehdä pakollista, valtakunnallista, aikatauluun sidottua veloitetta jokaiselle koululle ja opettajalle. (Saari 1983, 104.)

Koetoimintaa hoitamaan oli tutkimuslaitokseen perustettu vuonna 1971 erillinen Koulukoetoimisto, joka laajennuttuaan muutettiin vuoden 1973 alusta Evaluaatio-osastoksi. Huolimatta niin oppilaiden kuin opettajien myönteisestä suhtautumisesta koetoimintaan (Linnakylä 1974), jäi koulusaavutuskokeiden aika Suomessa kuitenkin melko lyhyeksi, kun kokeiden järjestäminen lakkautettiin vuonna 1976 toiminnan laajentamisen edellyttämän lisärahoituksen puutteessa (Laukkanen 1998). Näin peruskoulu-uudistuksen yhteydessä hahmotellut arvioinnin kehittämislinjaukset ”*järjestelmätasolla jatkuvasti hankittavasta tiedosta*” (KM 1970:A4, 158) tai ”*valtakunnallisista standardikokeista*” (KM 1973:38) jäivät käytännön tasolla toteutumatta.

Ensimmäinen laajempi oppimistulosten valtakunnallisesti edustava arviointi peruskoulussa toteutettiin vuonna 1979 äidinkielessä, matematiikassa ja englannin kielessä vuosiluokilla 4, 6 ja 9. Sen pääasiallisena tavoitteena oli selvittää, miten uudessa peruskoulujärjestelmässä koulutukselle asetetut tavoitteet toteutuivat ennen kaikkea alueellisella tasolla ja sukupuolten välillä. Seuraava samankaltainen, joskin useampia oppiaineita koskenut arviointi toteutettiin vasta vuonna 1990. Myös tällä tutkimuksella oli kohdennettu tavoite. *Peruskoulun arviointi 90* -tutkimuksen tehtävänä oli selvittää vuonna 1985 käyttöön otettujen opetussuunnitelman perusteiden toimivuus ja antaa tutkimustietoon pohjautuva vastaus julkisessa keskustelussa velloneisiin väitteisiin siitä, että tasokursien poistaminen olisi heikentänyt erityisesti lahjakkaiden oppilaiden oppimistuloksia ja -asenteita. Edelliseen arviointiin ja kansainvälisten tutkimusten tuloksiin verrattuna tälle väitteelle ei löytynyt tukea – päinvastoin peruskoulun todettiin olevan oppimistuloksiltaan julkista mainettaan parempi (Linnakylä & Saari 1993). Kolmas arviointi toteutettiin vuonna 1995 (Väljärvi 2008).

Nämä tilannekatsaustenkin nimellä kulkeneet ensimmäiset arviointikokeet toteutettiin Suomessa otospohjaisesti. Vielä tässä vaiheessa otantamenetelmän perustelussa ei viitattu koulukohtaisten tulosten julkistamiseen, vaan huolta kannettiin muun muassa arvioinnin kuormittavuudesta:

Huolestuneisuus runsaan mittaamisen vaikutuksista kouluihin, opettajiin ja oppilaisiin on synnyttänyt tarvetta vähentää oppilaiden testaamista. Koululaitosta koskevan tiedon hankkimiseen mahdollisimman taloudellisesti voidaan vaikuttaa mm. kehittämällä otantaratkaisuja. (Kupari 1980, III.)

Kokonaisuudessaan oppimistulosten kansallisen tason arviointia ennen 1990-lukua leimasivat kokonaisvaltaisen arviointistrategian puute, epäsystemaattisuus, arvioinnin kohdentuminen tiettyyn tarkoitukseen sekä niukka resursointi. Erilaisissa kehittämissuunnitelmissa mainitut tavoitteet tai visiot jäivät pitkälti toteuttamatta. (Laukkanen 1998; Männistö 1997; Välijärvi 2008.) Ei pidä ymmärtää, ettei kohdennettujen resurssien asettamisrajoissa olisi tehty kansainvälisestäkin korkealaatuista arviointitutkimusta (Välijärvi 2008). Voidaan kuitenkin ajatella, että arviointitoiminnan myöhäinen ja katkonainen institutionalisoituminen sekä tiedonkeruumenetelmän otospohjaisuus antoivat jo ennen 1990-lukua vahvan leimansa suomalaisen peruskoulujärjestelmän arviointikulttuurille.

1990-luvun koulutuspoliittiset murrokset – kohti nykymuotoista arviointitoimintaa Suomessa

1980- ja 1990-lukujen taitteessa toimeenpannut yhteiskuntapolitiiset uudistukset heijastuivat molempien maiden koulutuspolitiikkaan. Koulutuspolitiikan reformeja, kuten markkinoitumista, kilpailuideologian vahvistumista ja yksilöllisten valinnanmahdollisuuksien kasvattamista, on kuvattu tutkimuskirjallisuudessa laajasti (esim. Ahonen 2003; Lundahl 2002; Rinne 2000; Varjo 2007). Päätöksenteon desentralisaatio- ja deregulointipolitiikka edellytti arviointitoiminnalta entistä vahvempaa roolia koulutuksen seurannassa. Normiohjauksen purkaminen ja päätöksentekovallan siirtäminen paikallistasolle lisäsivät koulujen liikkumavaraa, ja kansallisen arvioinnin tehtäväksi vahvistui siten varmistaa, että koulutukselle asetetut tavoitteet toteutuisivat. Jälkikäteisestä tulosohjaukselle ominaisesta arvioinnista nousi näin 1990-luvun aikana neljäs

hallinnollinen koulutuksen ohjausmekanismi lainsäädännöllisen, taloudellisen ja ideologisen ohjauksen rinnalle (Lundgren 1991).

Suomessa vastuu kansallisen tason arvioinnista säädettiin vuonna 1991 perustetulle Opetushallitukselle. Päätös herätti keskustelua arvioinnin riippumattomuudesta ja viraston kaksoisroolista, kun sama organisaatio toimi näin sekä koulutuksen tavoitteiden määrittelijänä että oppimistulosten pääasiallisena arvioijana. Sittemmin 2000-luvulla seuranneet organisaatiomuutokset kytkeytyvätkin osin samaan arviointitiedon luotettavuuden vahvistamiseen (Lyytinen & Lukkarinen 2010).

Oppimistulosten mittaamisen painoarvo ja merkitys arviointitoiminnassa korostuivat entisestään, kun koulutarkastustoiminnasta luovuttiin 1990-luvun alussa. Opetushallituksen tehtäväksi muodostui siten määritellä suomalaisen perusopetuksen arviointitoiminnan strategiset tavoitteet ja käytännön toteuttamismuodot. Pekka Syrjäsen (2013) yksityiskohtaisesta analyysistä käy ilmi, että koko ikäluokan kattavien arviointikokeiden ja peruskoulun päätökkeiden mahdollisuus oli esillä Opetushallituksessa käydyissä keskusteluissa. Yhtäältä vaihtoehtoisia arviointitoiminnan muotoja käsiteltiin resurssikysymyksenä, mutta kuten edellä on todettu, nousi koulujen välisten ranking-listojen vastustaminen ehdottomasti tärkeimmäksi perusteluksi otospohjaisen arviointimallin esittämiselle. Vuosina 1994–1995 toimineen *Opetustoimen tuloksellisuuden arvioinnin metodisen kehittämisen* eli ARMI-projektin puheenjohtajan Ritva Jakku-Sihvosen johtama työryhmä päätyi esittämään otospohjaista oppimistulosten kansallista arviointia suomalaiseen koulujärjestelmään sopivaksi. Linjaus kirjattiin ensimmäisen kerran vuonna 1995 *Koulutuksen tuloksellisuuden arviointi* -raportissa (OPH 1995) ja uudelleen lainsäädäntötyön alla 1998 (OPH 1998a; 1998b). Nykymuotoiset arviointikäytännöt saivat lainvoimaisen vahvistuksen vuonna 1999 voimaan astuneessa perusopetuslaissa (L 628/1998), jossa muotoiltiin, että ainoastaan ”arviointien keskeiset tulokset tulee julkistaa” (§ 21)¹⁴. Otanta-

¹⁴ Sivistysvaliokunnan mietinnössä annettiin laintulkinnalle eksplisiittisempi tarkennus: ”(...) lainsäädännön tarkoituksena [ei] näin ollen ole julkistaa yksittäistä koulua tai yksittäistä opettajaa koskevia tietoja. Arviointitulosten julkaiseminen ei missään tapauksessa saa johtaa koulujen

menetelmästä voidaan sanoa muodostuneen näin eräänlainen *tiedonkeruutekninen portinvartija* koulujen välisen julkisen vertailtavuuden estämiseksi.

2000-luvun alussa ovat median edustajat kahteen otteeseen oikeusteilse vaatineet koulukohtaisia tietoja kuntatason arvioinneista julkisuuslakiin vedoten (Simola 2005). Siitäkin huolimatta, että toisessa tapauksista korkein hallinto-oikeus velvoitti kyseisen kunnan luovuttamaan osan tiedoista, ei suurta painetta arviointitulosten julkistamiseen ole Suomessa esiintynyt. Suomalaisen arviointimallin periaatteet (ks. Simola, Varjo & Rinne 2010) näyttäisivätkin nauttivan poikkeuksellisen laajasta ja konsensuksenomaisesta hyväksynnästä suomalaisessa yhteiskunnassa, jota PISA-menestyksen kontingenssi on entisestään tukenut. Tosin viimeisimpien vuosien havainnot ovat kertoneet oppimistulosten yleisestä laskusuunnasta myös Suomessa. Tämän kehityksen mahdollisiin merkityksiin arviointipolitiikalle palataan luvun pohdintaosiossa.

Oppimistulosten arviointi pakolliseksi ja tulosten julkistaminen Ruotsissa

1990-luvun koulureformit toteutuivat Ruotsissa Suomeen verrattuna huomattavasti radikaalimmin. Julkisesti rahoitettujen yksityiskoulujen (*friskola*) lukumäärä ja suosio kasvoivat nopeasti niin sanotun *voucher*-käytännön käyttöönoton myötä vuonna 1992. Vanhempien aiempaa vapaampi kouluvalinta sai uudessa kvasimarkkinamallissa siten korostetumman merkityksen kuin Suomessa. (Lundahl 2002). Vuonna 1991 perustettu koulutuksen kansallinen keskusvirasto Skolverket jatkoi lakkautetun Skolöverstyrelsénin työtä oppimistulosten arvioinnissa.

paremmuusjärjestykseen asettamiseen eikä leimata kouluja, opettajia ja oppilaita tasoltaan heikoiksi ja hyviksi yksipuolisin perustein.” (SiVM 3/1998.) Samassa mietinnössä tosin mainitaan, että arviointien tulisi *lisäksi* tuottaa tietoa opintoja koskevien valintojen perusteeksi opiskelijoille ja heidän huoltajilleen (SiVM 3/1998), mutta tämän merkitys käytännössä on jäänyt epäselväksi. Ensisijaisesti suomalainen arviointitoiminta ja sen tulokset on kohdistettu koulutuksen päätöksentekijöiden käyttöön koulutuksen kehittämiseksi. Otoksessa mukana olleille kouluille toimitetaan kokeista koulukohtaiset palautetiedot niin, että omien oppilaidensa tuloksien lisäksi koulu saa tiedot kansallisista keskiarvoista (Jakku-Sihvonen 2013, 25).

Markkinalogiikasta vaikutteita saaneet koulupolitiikan reformit heijastuivat 1990-luvun aikana myös koulutuksen arviointipolitiikkaan (ks. Lundahl 2009). Vuoden 1994 opetussuunnitelman uudistuksen myötä Ruotsissa siirryttiin kriteeriperustaiseen oppilasarviointiin. Yksityiskohtaisesti määriteltyjen arvosanakriteerien keskeisenä tavoitteena oli edelleen parantaa arvosanamennettelyn yhdenvertaisuutta ja ratkaista suhteelliseen oppilasarviointiin sisällyneet ongelmat. Samalla aiempaa tarkemmin määriteltyt oppimisen kriteerit mahdollistivat valtiotasolla tarkemman koulujärjestelmän seurannan. Kouluille ja opettajille 53 vuoden ajan vapaaehtoinen standardikoe lakkautettiin ja korvattiin lukuvuodesta 1997–1998 lähtien koko ikäluokan pakollisella kansallisella kokeella (*nationell prov*). Pakollisuus koski aluksi kaikkien koulujen yhdeksännen vuosiluokan oppilaita, ja koe järjestettiin äidinkielessä, matematiikassa ja englannin kielessä – viidennellä vuosiluokalla vastaava koe säilyi vielä tuolloin vapaaehtoisena (Eurydice 2009b).

Tämän tutkimuksen kannalta toinen merkittävä uudistus koskee arviointikokeiden tulosten julkistamista. Koulujen oppimistulokset tuotiin pian koko ikäluokan pakollisten arviointikokeiden käyttöönoton jälkeen näkyvämmiin kansalaisten tietoisuuteen. Vuodesta 2001 lähtien koulukohtaiset arviointikokeiden tulokset on koottu Skolverketin internetsivuille SIRIS- ja SALSA-tietokantoihin kaikkien vapaasti saataville. Tietojen julkistaminen ei ole jäänyt koskemaan ainoastaan arviointikokeita vaan kattaa laajasti koko perusopetuksen ja lukion arviointitoiminnan. Esimerkiksi, 1990-luvun alkuvuosina rajusti supistettu koulutarkastustoiminta otettiin uudelleen käyttöön entistä vahvempana vuonna 2006, ja myös nämä tarkastusraportit ovat nykyään pääosin julkisia. Vaikka Skolverket ei itse ole esittänyt kouluja ”paremmuusjärjestyksessä”, on tiedon julkisuus mahdollistanut ruotsalaisen median koostaa monenlaisia ranking-listoja kouluista. Myös yksittäiset koulut ovat voineet käyttää arviointitietoja markkinoinnissaan ja kilpaillessaan uusista oppilaista.

2000-luvulla oppimistulosten kansallista arviointia on Ruotsissa edelleen tehostettu. Vuodesta 2006 oikeistohallitus on lisännyt sekä arviointikokeiden lukumäärää että arvioitavia oppiaineita.

Kokeiden pakollisuus on laajentunut koskemaan myös alakoulun oppilaita, ja kansallisia kokeita on järjestetty vuosiluokilla 3, 5 ja 9 aluksi äidinkielessä, matematiikassa ja englannin kielessä, sittemmin myös muissa oppiaineissa (Seegerholm 2009). Viime vuosina kritiikki ”liiallista mittaamista” kohtaan on voimistunut, ja muun muassa Ruotsin opettajien ammattijärjestöt ovat vaatineet kansallisten kokeiden määrän supistamista (Svenska Dagbladet 2014).

Voidaan ajatella, että aiemmin taulukossa 1 kuvatut kansallisen arvioinnin moninaiset tavoitteet näkyvät Ruotsin nykyisissä arviointikäytännöissä. Koko ikäluokan pakolliset arviointikokeet tuottavat kullekin koululle ja opettajalle entistä tarkempaa tietoa oman koulun tai luokan osaamistasosta suhteessa koko maan osaamiseen. Vaikka kouluilla on säilynyt oikeus päättää itse kansallisten kokeiden tulosten painoarvo omassa oppilasarvioinnissaan, on niiden ajateltu tukevan opettajien oppilasarviointia ja edistävän siten ennen kaikkea peruskoulun päättävien oppilaiden arvosanakohTELUN yhdenvertaisuutta. (Eurydice 2009b; Ouakrim-Soivio 2013, 56–63.)

Yhdenvertaisuuden edistämisen ohella muutokset ovat sisältäneet myös julkilausuttuja järjestelmätason kontrollin muotoja. Ruotsin opetusministeriö on esittänyt raportissaan, kuinka arviointitoiminnan tehostamiset olivat looginen seuraus ruotsalaiskoulujen rajusti heikentyneistä oppimistuloksista. Raportin mukaan jo vuodesta 1995 lähtien järjestelmätasolla havaitut eroavaisuudet eri kuntien ja koulujen yleisestikin laskeneissa oppimistuloksissa edellyttivät kansallisen otteen selvää tiukentamista. Alun perin kansallisesti havaitut ongelmat saivat lopulta PISA-tutkimusten myötä kansainvälistä lisäevidenssiä perusopetuksen heikosta tasosta, mikä seurauksena arviointitoimintaa tiukennettiin entisestään systematisoimalla koulutarkastustoimintaa ja lisäämällä kansallisten arviointitestien lukumäärää ja oppiaineita. (ME 2010, 5.)

Tulokset – arviointipolitiikan alkujuuret nykykäytäntöjen mahdollistajina

Luvun tutkimuskysymyksenä esitettiin, miten varhaisemmat oppimistulosten arvioinnin tavoitteet ja menetelmät ovat mahdollistaneet nykymuotoisten arviointikäytäntöjen eroavaisuudet Suomessa ja Ruotsissa. Historiallisen institutionalismin viitekehyksessä pyrittiin havaitsemaan molemmissa maissa polkuriippuvuuksia, jotka ovat omalta osaltaan mahdollistaneet myöhemmin toteutuneet kehityskulut.

Keskeisin havainto koskee arviointitoiminnan *institutionaalista kestoa*. Maiden arviointitoiminnan juuret ovat eri aikakauden ja yhteiskunnallisen kehitysvaiheen tuotetta. Ruotsia voidaan pitää yhtenä Euroopan, ellei koko maailman pioneerimaana oppimistulosten arvioinnissa 1900-luvun alkupuoliskolla. Ajan hengen mukaisesti kvantitatiivinen psykometriikkaan perustuva mittaaminen koettiin edistyskelliseksi, ja sen varaan asetettiin suuria toiveita myös laajemmin yhteiskuntasuunnittelussa. Suomessa kansallinen arviointi sai alkunsa vasta 1970-luvulla peruskoulu-uudistuksen yhteydessä, 30 vuotta Ruotsia myöhemmin. Voidaan siis ajatella, että 1990-luvun arviointilinjauksia edelsivät eripituiset institutionaaliset perinteet, jotka ovat asettaneet erilaisia rajoitteita toimijoiden päätöksentekoon.

Toisena havaintona nostetaan esiin kansallisen arviointikokeen *käytön laajuus*. Ruotsissa 1940-luvulla käyttöönotetun standardikokeen pääasiallinen funktio oli ennen kaikkea edistää oikeudenmukaisen oppilasarvioinnin toteutumista koko maan kouluissa. Tämä tavoite edellytti riittävän laajamittaista kokeen käyttöä sekä systemaattista ja toistuvaa oppimistulosten mittaamista. Vapaaehtoinen koe saikin ruotsalaisilta opettajilta heti alusta hyvän vastaanoton. Kokeen suosio säilyi suurena koko sen reilun 50 vuoden historian ajan. Suomessa ”kansallisten kokeiden” käyttö oli huomattavasti rajatumpaa ja koski vain murto-osaa opettajakunnasta ja oppilaista. Ruotsissa valtaosalla niin kouluista, opettajista kuin oppilaistakin oli omakohtaista kokemusta kansallisen kokeen käytöstä. Voidaan siten ajatella, että kokeen käytön kokemukset ovat

kertautuneet tuleville opettajille ja tiiviiksi osaksi ruotsalaista koulutyötä eri laajuudella kuin Suomessa.

Kolmantena erona maiden välillä korostetaan kansallisen kokeen käytön ja hyödyntämisen *systemaattisuutta*. Jo alusta lähtien Ruotsissa oli selvää, että standardikokeella tavoiteltu oppilasarvioinnin yhdenvertaistuminen edellytti kokeen systemaattista käyttöä. Vaikka opettajat ja välillisesti myös oppilaat mainitaan yleensä kokeen käytön tärkeimpinä hyötyjinä, hyötyivät myös monet koulutuksen arviointitutkijat tutkimusalansa painoarvon kasvusta. Myös analysoitavan datan määrän nopea kasvu ja ennen kaikkea toistettavuus laajensi muun muassa erilaisten seuranta-asetelmien mahdollisuuksia. Tällä ei tarkoiteta, ettei myös Suomessa olisi tehty laadukasta arviointitutkimusta. Niukat resurssit ja tietyn koulutuspoliittisen keskustelun ympärille kietoutuneet tilannekatsaukset 10 vuoden välein kertovat pikemmin omaa kieltään päätöksentekijöiden suhtautumisesta systemaattisemman arvioinnin hyödyistä Suomessa. Ilmeistä on, että kokonaisvaltaisemman oppimistulosten arvioinnin hyödyistä puhuminen on jäänyt Suomessa eri aikakausina vaihtuvien vasta-argumenttien jalkoihin.

Neljäs yleisempi huomio molemmista maista koskee *arviointitiedon julkisuuden merkityksen muutosta*. Oppimistulosten rajaaminen opettajien, koulujen ja arviointitutkijoiden käyttöön miellettiin alkuvaiheessa itsestäänselvyydeksi, eikä arviointitiedon julkisuuteen ole osattu juurikaan kohdistaa ulkopuolisia vaatimuksia ennen 1990-lukua. Tulosten julkistamisen problematiikka nousee tärkeämpään osaan vasta kouluvalinnan vapauttamisen ja ranking-kulttuurin leviämisen myötä. Siltikin, tulosten julkisuus itsessään saa maissa erilaisen merkityksen. Yksityiskohtaisen arviointitiedon julkistaminen ei ole missään vaiheessa noussut Ruotsissa samanlaiseen päätöksenteon rooliin kuin Suomessa. Kun Suomessa tiedon julkistamisen haittavaikutukset ovat vaikuttaneet vahvasti itse tiedonkeruumenetelmän valintaan, on Ruotsissa prosessi ollut päinvastainen – päätös koko ikäluokan mittaamisesta johti julkisuusperiaatteen myötä nopeasti koulukohtaisten tietojen vapaaseen saatavuuteen. Näyttäisi siis siltä, että koulujen oppimistulosten potentiaalinen käyttäjäkunta ymmärretään jo lähtökohtaisesti eri tavoin.

Yhteenvetona voidaan siis todeta, että maiden arviointipolitiikan linjaukset ja käytännöt 1990-luvulla ovat yhtäältä tulkittavissa harkituiksi ja johdonmukaisiksi poliittisiksi päätöksiksi, joihin vaikutti yleinen käsitys maan peruskoulujärjestelmän tilasta. Toisaalta, historiallisen institutionalismin valossa voidaan ajatella, että Ruotsi ja Suomi myös vastasivat 1990-luvulla esiin nousseisiin kysymyksiin varsin erilaisista lähtökohdista ja arvioinnin traditioista. Piersonin (2000) mukaan historiallisen institutionalismin ja polkuriippuvuuksien tutkimisessa on tärkeää esittää hypoteeseja siitä, olisiko jokin vaihtoehtoinen kehityskulku ollut todennäköinen ilman tiettyä historiallista kehitystä. Edellä esitetyn perusteella voidaan esittää, että arviointitoiminnan varhaisemmilla tavoitteilla ja käytännöillä on polkuriippuvuuksiksi tulkittavissa olevia historiallisia vaikutuksia. Voidaan yhtäältä pitää epätodennäköisenä, että Ruotsin tilivelvollisuutta ja valtion tiukempaa kontrollia korostavat nykyarviointikäytännöt – koko ikäluokan pakollisen arvioinnin laajentaminen ja koulukohtaisten tulosten julkistaminen – olisivat toteutuneet sellaisinaan ilman maan pitkää historiallista traditiota standardikokeen systemaattisessa ja myönteisessä käytössä. Toisaalta, voidaan esittää, että koska koulutuksen arviointi ei ehtinyt Suomessa saada Ruotsin kaltaista vakiintunutta ja systemaattiseen mittaamiseen perustuvaa strategista muotoa eikä koko ikäluokan kattavaa arviointijärjestelmää ennen 1990-lukua, oli oppimistulosten julkistamisen torjuminen helpompaa siihen liitettyjen kielteisten haittavaikutusten nojalla. Nykymuotoinen arviointitoiminta jatkoi näin ollen otospohjaisten kansallisten arviointien toteuttamisen perinnettä.

Pohdinta

Tässä tutkimuksessa vertailtiin peruskoulun oppimistulosten kansallisen tason arviointia kahdessa Pohjoismaassa, Suomessa ja Ruotsissa. Luvussa sivuttiin myös laajemmin arviointikäytäntöjä kaikissa Pohjoismaissa, joiden keskuudessa Suomi näyttäytyi poikkeustapauksena. Jo pelkästään tämä havainto herättää kysymyksiä

siitä, voidaanko ylipäänsä puhua yhtenäisestä pohjoismaisesta arviointikulttuurista.

Lienee syytä vielä selventää, ettei kirjoituksessa ole ollut tarkoitus väheksyä päätöksentekijöiden tärkeää ja vastuullista roolia arviointitoiminnan linjaajina vaan pikemmin pohtia toiminnan mahdollisuuksia ja rajoitteita. Suomessa arviointikäytäntöjen myöhäinen, epäsystemaattinen ja katkonainen institutionaalistuminen kääntyi alun perin tarkoittamattomasti nykymuotoisen kehittävän arvioinnin tunnuspiirteiden eduksi. Otospohjaisesta arviointimenetelmästä, joka aikanaan syntyi kokeilunomaiseen ja kohdennettuun arviointitarpeeseen, muodostui 1990-luvulla oppimistulosten julkistamiskysymyksen ratkaiseva tekninen portinvartijamekanismi, joka on asettanut tarkat rajat arviointi-informaation käyttäjäkunnalle.

Historiallisten polkuriippuvuuksien tutkimisen ohella tutkimusaihetta olisi tärkeä lähestyä myös muista näkökulmista. Esimerkiksi sosiologisen ja diskursiivisen institutionalismin selitysmallien hyödyntäminen voisi antaa vielä kattavamman kuvan maiden välisiä eroja selitettäessä. Muun muassa arviointitiedon luonne, viranomaistiedon avoimuus ja ylipäänsä koulun tai opettajan ammatillinen status yhteiskunnassa ja päätöksenteossa ovat tekijöitä, joilla lienee oma merkityksensä Suomen ja Ruotsin arviointikulttuurin eroja selitettäessä. Vertailevaan tutkimukseen riittää siten saman teeman ympärille monia lisätutkimuksen mahdollistavia tulokulmia.

Luvun päätteeksi pohditaan arviointipolitiikan muutospaineita kahden ajankohtaisen koulutuspoliittisen keskustelun kautta. Suomessa viimeisimmät PISA-tulokset avasivat keskustelua maamme peruskoulun tasosta. PISA 2012 (Kupari, Välijärvi, Andersson, Arffman, Nissinen, Puhakka & Vettenranta 2013) osoitti muiden tutkimusten (mm. Bernelius 2013, Kupiainen, Marjanen, Vainikainen & Hautamäki 2011) tavoin oppimistulosten heikentyneen ja koulujen tulosten erojen kasvaneen. Varsin pian tulosten julkistamisen jälkeen opetusministeriön kokoama Tulevaisuuden peruskoulu -kehittämishanke kokosi maan koulutusasiantuntijat yhteen pohtimaan keinoja laskusuunnan pysäyttämiseen (OKM 2014).

Voidaanko odottaa, että oppimistulosten lasku johtaisi Suomessa Ruotsin tavoin tiukempaan kansallisen tason kontrolliin ja arviointikäytäntöjen muuttamiseen? Tämän tutkimuksen pohjalta voidaan todeta, ettei ulkoisen arvioinnin lisäämiselle tai ainaakaan koko ikäluokan arviointijärjestelmän perustamiselle näyttäisi olevan kovinkaan vankkoja historiallisia tai kulttuurisia perusteita. Siitäkin huolimatta, että myös Suomessa monissa tutkimuksissa on havaittu opettajien antamien arvosanojen ja arviointikokeissa todennetun todellisen osaamistason välillä systemaattista vaihtelua (esim. Kuusela 2006; Ouakrim-Soivio 2013), ei oppimistulosten mittaaminen koko ikäluokan yhteneväisillä kokeilla ole saanut laajaa kannatusta. Aika ajoin arvosanamenettelyn yhdenvertaisuus nousee Suomessa julkiseen keskusteluun. Esimerkkinä tästä on kasvatopsykologian professorin Liisa Keltikangas-Järvisen kommentti, jossa hän muistutti oppilasarvioinnin ongelmista Suomessa viitaten samalla muiden Pohjoismaiden koko ikäluokan arviointiin (Helsingin Sanomat 6.5.2014). Oletettavaa on, että vaihtoehtoisten arviointikäytäntöjen avaukset törmäävät jatkossakin laajaan vastustukseen, jossa viitataan ranking-listojen haittoihin.

Ruotsissa vuoden 2014 parlamenttivaalit päättyivät sosialidemokraattien voittoon kahden oikeistohallituskauden jälkeen. Koulutus oli vaalien alla yksi keskeisistä vaaliteemoista, ja kritiikki oikeiston ajamaa koulutuspolitiikkaa kohtaan on ollut vahvaa. Voidaanko olettaa, että valtasuhteiden vaihtuminen johtaisi esimerkiksi tilanteeseen, jossa koulukohtaisten oppimistulosten julkistamisesta luovuttaisiin? Kirjoituksessa esitetyn pohjalta tällainen muutos on epätodennäköinen. Tulosten julkistaminen ei ole ollut samanlainen kynnyskysymys kuin Suomessa. Koska arviointitiedon julkisuus on ensi kädessä artikuloitu kansalaisten demokraattiseksi oikeudeksi, voitaisiin muutos tulkita demokraattisten oikeuksien kaventamiseksi – mikä ei Ruotsin kaltaisessa maassa tunnu kovinkaan todelliselta vaihtoehdolta. Tässä mielessä poliittisten päätösten kumoaminen voi olla vaikeaa, vaikka ne osoittautuisivatkin myöhemmin haitallisiksi. Oletettavampaa sen sijaan on, että arviointikokeiden lukumäärää saatetaan kritiikin myötä asteittain vähentää.

Voidaan kuitenkin arvella, että koulujen toimintaan kohdistuu Suomessa jatkossa painokkaampia läpinäkyvyyden vaatimuksia, mikäli oppimistulosten heikentymissuunta jatkuu ja etenkin jos koulujen eriytyminen kiihtyy voimakkaasti. Nähtäväksi jää, minkälaisen roolin oppimistulosten kansalliset arviointilinjaukset tällöin keskustelussa saavat. Mikäli julkistamispainetta kasvaisi radikaalisti, olisi olennaista tällöin etsiä aktiivisesti keinoja, joilla tietojen julkistamisen haittavaikutukset saataisiin minimoitua. Samalla on kuitenkin aiheellista muistaa kysyä, korostaako vertailtavassa muodossa esitettävä arviointitieto itsessään aina enemmän koulujen välisiä suhteellisia eroavaisuuksia kuin niiden suhteellista tasa-arvoisuutta.

Lähteet

- Ahonen, S. 2003. Yhteinen koulu – tasa-arvoa vai tasapäisyyttä? Koulutuksellinen tasa-arvo Snellmanista tähän päivään. Tampere: Vastapaino.
- Antikainen, A. 2006. In search of the Nordic model in education. *Scandinavian Journal of Educational Research* 50 (3), 229–243.
- Bernelius, V. 2013. Eriytyvät kaupunkikoulut. Helsingin peruskoulujen oppilaspohjan erot, perheiden kouluvalinnat ja oppimistuloksiin liittyvät aluevaikutukset osana kaupungin eriytymiskehitystä. Helsingin kaupungin tietokeskus, tutkimuksia 1.
- Björklund, A., Clark, M., Edin, P.-A., Fredriksson, P. & Krueger, A. 2005. The market comes to education in Sweden: An evaluation of Sweden's surprising school reforms. New York: Russell Sage Foundation.
- Department for Education. 1994. Our children's education: the updated parent's charter. London, DfE.
- Erkkilä, T. 2010. Reinventing Nordic openness: Transparency and state information in Finland. Helsinki: Helsingin yliopisto, Acta politica, no. 40.
- Eurydice 2009a. National testing of pupils in Europe: Objectives, organisation and use of results. EACEA: Eurydice. Bryssel. DOI: 10.2797/18294.
- Eurydice 2009b. National testing of pupils in Europe: Objectives, organisation and use of results. Sweden. EACEA: Eurydice. Bryssel.
- Fredriksson, V. (toim.) 1950. Svenska folkskolans historia: 5. delen: Det svenska folkundervisningsväsendet 1920–1942. Stockholm: Albert Bonniers förlag.
- Helsingin Sanomat. 6.5.2014. Marjo Valtavaara: Tutkijat: Ujous laskee poikien arvosanoja. <http://www.hs.fi/kotimaa/Tutkijat+Ujous+laskee+poiki+en+arvosanoja/a1399263864827>. (Luettu 6.5.2014.)
- Henricson, S.-E. 1987. Skolöverstyrelsens provverksamhet – en översikt 1965–1985. Rapport R 87:3. Stockholm: Skolöverstyrelsen.
- Henrysson, S. 1969. Educational testing in Sweden. Teoksessa K. Ingenkamp

- (toim.) Developments in educational testing, vol. 1. London: University of London Press, 80–86.
- Hudson, C. 2007. Governing the governance of education: the state strikes back? *European Educational Research Journal* 6 (3), 266–282.
- Jakku-Sihvonen, R. 2010. Oppimistulosten arviointitiedon puntarointia. *Hal-
linnon Tutkimus* 29 (4), 317–324.
- Jakku-Sihvonen, R. 2013. Oppimistulosten arviointijärjestelmistä ja niiden kehittämishaasteista. Teoksessa A. Räisänen (toim.) *Oppimisen arvi-
oinnin kontekstit ja käytännöt. Opetushallituksen raportit ja selvitykset* 2013:3. Helsinki: Opetushallitus, 13–36.
- Karsten, S., Visscher, A. & de Jong, T. 2001. Another side to the coin: the unintended effects of the publication of school performance data in Eng-
land and France. *Comparative Education* 37 (2), 231–242.
- KM 1970: A4. Peruskoulun opetussuunnitelmakomitean mietintö I. Komi-
teamietintö 1970: A4. Helsinki: Valtioneuvosto.
- KM 1973: 38. Oppilasarvostelun uudistamistoimikunnan mietintö. Komitea-
mietintö 1973: 38. Helsinki: Valtioneuvosto.
- Kupari, P. 1980. Peruskoulun matematiikan opetuksen tilannekartoituksen
suunnittelu ja toteutus. Jyväskylän yliopisto. Kasvatustieteiden tutkimus-
laitoksen julkaisuja 306/1980.
- Kupari, P., Välijärvi, J., Andersson, L., Arffman, I., Nissinen, K., Puhakka, E.
& Vettenranta, J. 2013. PISA 2012 ensituloksia. Opetus- ja kulttuurimi-
nisteriön julkaisuja 2013:20. Opetus- ja kulttuuriministeriö & Koulutuk-
sen tutkimuslaitos.
- Kupiainen, S., Marjanen, J., Vainikainen, M.-P. & Hautamäki, J. 2011. Oppi-
maan oppiminen Vantaan peruskouluissa. Kolmas-, kuudes- ja yhdeksäs-
luokkalaiset oppijoina keväällä 2010. Vantaan kaupungin sivistysvirasto
ja Helsingin yliopiston Koulutuksen arviointikeskus.
- Kuusela, J. 2006. Temaattisia näkökulmia perusopetuksen tasa-arvoon.
Opetushallitus. Oppimistulosten arviointi 6/2006.
- L 628/21.8.1998. Perusopetuslaki.
- Laukkanen, R. 1998. Opetustoimen evaluaatioajattelun kehitys Suomessa
1970-luvulta 1990-luvulle. Jyväskylän yliopisto. Koulutuksen tutkimuslai-
tos. Tutkimuksia 5.
- Leimu, K. (toim.) 2004. Kansainväliset IEA-tutkimukset Suomi-kuvaa luo-
massa. Jyväskylä: Koulutuksen tutkimuslaitos.
- Linnakylä, P. 1974. Oppilaiden ja opettajien asenteet yhteisiä kokeita
kohtaan sekä kokemukset ja odotukset yhteisten kokeiden funktiois-
ta. Jyväskylän yliopisto. Kasvatustieteiden tutkimuslaitoksen julkaisuja
222/1974.
- Linnakylä, P. & Saari, H. 1993. Oppiiko oppilas peruskoulussa? Peruskoulun
arviointi 90 -tutkimuksen tuloksia. Jyväskylän yliopisto. Kasvatustietei-
den tutkimuslaitos.
- Ljung, B.-O. 2000. Standardproven – 53 år i skolans tjänst. Rapport från
PRIM-gruppen nr. 17. Stockholm: Lärarhögskolan i Stockholm, Institutio-
nen för undervisningsprocesser, kommunikation och lärande.

- Lundahl, C. 2009. Varför nationella prov? – framväxt, dilemma, möjligheter. Lund: Studentlitteratur.
- Lundahl, L. 2002. Sweden: decentralization, deregulation, quasi-markets – and then what? *Journal of Education Policy* 17 (6), 687–697.
- Lundgren, U. 1991. Educational policymaking, decentralization and evaluation. Teoksessa M. Granheim, M. Kogan & U. P. Lundgren (toim.) *Evaluation as policymaking. Introducing evaluation into a national decentralized educational system*. London: Jessica Kingsley Publishers, 66–88.
- Lyytinen, H. K. & Lukkarinen, E. 2010. Arvioinnin lakia niin kuin sitä luetaan historiasta nykypäivään. Teoksessa E. Korkeakoski & P. Tynjälä (toim.) *Hyötyä ja vaikuttavuutta arvioinnista. Koulutuksen arviointineuvoston julkaisuja 50*. Jyväskylä: Koulutuksen arviointineuvosto, 77–98.
- Mahoney, J. 2000. Path dependence in historical sociology. *Theory and Society* 29, 507–548.
- Maw, J. 1999. League tables and the press – value added? *Curriculum Journal* 10 (1), 3–10.
- ME. 2010. OECD review on evaluation and assessment frameworks for improving school outcomes: Country background report for Sweden. Ministry of Education and Research Sweden.
- Männistö, Y. 1997. Kouluhallitus koulututkimuksen rahoittajana ja tutkimustiedon käyttäjänä. Helsingin yliopiston opettajankoulutuslaitoksen tutkimuksia 181.
- OECD. 2011. OECD reviews of evaluation and assessment in education – Norway. <http://www.oecd.org/norway/48632032.pdf>. (Luettu 6.5.2014.)
- OKM. 2014. Tulevaisuuden peruskoulu -kehittämishanke. http://www.minedu.fi/OPM/Koulutus/koulutuspolitiikka/Hankkeet/tulevaisuuden_peruskoulu/. (Luettu 1.10.2014.)
- OPH. 1995. Koulutuksen tuloksellisuuden arviointimalli. Arviointi 9/1995. Helsinki: Opetushallitus.
- OPH. 1998a. Kansallinen oppimistulosten arviointijärjestelmä. Arviointi 4/1998. Helsinki: Opetushallitus.
- OPH. 1998b. Koulutuksen tuloksellisuuden arviointimalli. Arviointi 7/1998. Helsinki: Opetushallitus.
- Ouakrim-Soivio, N. 2013. Toimivatko päättöarvioinnin kriteerit? Oppilaiden saamat arvosanat ja Opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiserojen mittareina. Helsinki: Opetushallitus.
- Ouakrim-Soivio, N., Rinkinen, A. & Karjalainen, T. (toim.) 2015. Tulevaisuuden peruskoulu. Opetus- ja kulttuuriministeriön julkaisuja 8:2015. Opetus- ja kulttuuriministeriö.
- Ozga, J., Dahler-Larsen, P., Segerholm, C. & Simola, H. (toim.) 2011. *Fabricating quality in education: Data and governance in Europe*. New York: Routledge.
- van Petegem, P., Vanhoof, J., Daems, F. & Mahieu, P. 2005. Publishing information on individual schools? *Educational Research and Evaluation* 11 (1), 45–60.
- Pierson, P. 2000. The limits of design: Explaining institutional origins and

- change. *Governance* 13 (4), 475–499.
- Power, S. & Frandji, D. 2009. Education markets, the new politics of recognition and the increasing fatalism towards inequality. *Journal of Education Policy* 25 (3), 385–396.
- Rinne, R. 2000. The globalisation of education: Finnish education on the doorstep of the new EU millennium. *Educational Review* 52 (2), 131–142.
- Rinne, R., Kivirauma, J. & Lehtinen, E. 2004. *Johdatus kasvatustieteisiin*. 5. uudistettu painos. Helsinki: WSOY.
- Rönnberg, L., Lindgren, J. & Segerholm, C. 2013. In the public eye: Swedish school inspection and local newspapers: exploring the audit–media relationship. *Journal of Education Policy* 28 (2), 178–197.
- Saari, H. 1983. Koulukoetoiminnan periaatteista peruskoulussa. Teoksessa V. Hirvi (toim.) *Peruskoulun kehittäminen tutkimustulosten perusteella*. Tutkijoiden artikkeleita peruskoulusta ja sen kehittämisestä. Jyväskylän yliopisto. Kasvatustieteiden tutkimuslaitos. Selosteita ja tiedotteita 209/1983, 104–5.
- Sahlberg, P. 2011. *Finnish lessons: What can the world learn from educational change in Finland?* New York: Teachers College Press.
- Segerholm, C. 2009. 'We are doing well on QAE': the case of Sweden. *Journal of Education Policy* 24 (2), 195–209.
- Seppänen, P. 2006. Koululavalintapolitiikka perusopetuksessa – suomalaiskaupunkien koulumarkkinat kansainvälisessä valossa. *Kasvatusalan tutkimuksia* 26. Turku: Suomen kasvatustieteellinen seura.
- Shiel, G., Kellaghan, T. & Moran, G. 2010. *Standardised testing in lower secondary education*. Dublin: National Council for Curriculum and Assessment, Research Report no. 12.
- Simola, H. 2005. Koulukohtaiset oppimistulokset ja julkisuus. *Yhteiskuntapolitiikka* 70 (2), 179–187.
- Simola, H., Varjo, J. & Rinne, R. 2010. Vasten valtavirtaa – kontingenssi, polkuriippuvuus ja konvergenssi suomalaisen perusopetuksen laadunarviointimallin kehityskuluissa. *Hallinnon tutkimus* 29 (4), 285–302.
- SiVM 3/1998. Sivistysvaliokunnan mietintö SiVM 3/1998. Hallituksen esitys koulutusta koskevaksi lainsäädännöksi.
- Skolverket. "SIRIS a source of information for understanding of the world of education". <http://siris.skolverket.se/siris/f?p=SIRIS:33:0>. (Luettu 1.10.2014.)
- SOU 1938:29. *Intagning av elever i första klassen av de allmänna läroverken*. Stockholm: 1938.
- SOU 1945:45. *Skolans betygsättning, 1940 års skolutrednings betänkanden och utredningar*, V. Stockholm: 1945.
- Svenska Dagbladet. 28.3.2014. Alldeles för många nationella prov. SvD opinion: Brännpunkt.
- Syrjänen, P. 2013. Uutta perinnettä rakentamassa – Suomen peruskoulujen otantapohjaisen arviointimallin synty politiikan dynamiikan näkökulmasta. Pro gradu -tutkielma. Helsingin yliopisto. Käyttätymistieteiden laitos.
- Telhaug, A. O., Mediås, O. A. & Aasen, P. 2006. The Nordic model in

- education: Education as part of the political system in the last 50 years. *Scandinavian Journal of Educational Research* 50 (3), 245–283.
- Thelen, K. 1999. Historical institutionalism in comparative politics. *Annual Review of Political Science* 2, 369–404.
- Varjo, J. 2007. Kilpailukykyvaltion koululainsäädännön rakentuminen. Suomen eduskunta ja 1990-luvun koulutuspoliittinen käänne. Helsingin yliopisto. Kasvatustieteen laitoksen tutkimuksia 209.
- Visscher, A. J. 2001. Public school performance indicators: Problems and recommendations. *Studies in Educational Evaluation*, 27 (2001), 199–214.
- Välijärvi, J. 2008. Arviointi – itsetarkoitus vai pedagogiikan työkalu? Artikkelisarjassa R. Meriläinen (toim.) *Suomalaisen koulutuspolitiikan murros 1990-luvulla*. Helsinki: OKKA-Säätiö, 32–46.
- Wallenius, T. 2015. Justifying opposite publication policies of school performance results in Finland and Sweden. Teoksessa S. Jokila, J. Kallo & R. Rinne (toim.) *Comparing times and spaces: Historical, theoretical and methodological approaches to comparative education*. Kasvatusalan tutkimuksia 69. Jyväskylä: Suomen kasvatustieteellinen seura, 209–231.
- West, A. & Pennell, H. 2000. Publishing school examination results in England: incentives and consequences. *Educational Studies* 26 (4), 423–436.

Schools, accountability and transparency—approaching the Nordic school evaluation practices through discursive institutionalism

Tommi Wallenius , Sara Juvonen , Petteri Hansen  and Janne Varjo 

Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

ABSTRACT

Over recent decades we have witnessed a growing emphasis on educational quality assurance and evaluation (QAE) around the globe. The trend, not only to intensify evaluative measurements, but also to publish school-specific indicators, has become visible also in the Nordic countries. In Sweden, Denmark and Norway, the governments have launched web-portals, in which various indicators can be observed and compared at the school level. However in Finland, the data is published only at a general level.

In this article we compare the discourses of educational experts on comprehensive school QAE policies and practices in four Nordic countries, Sweden, Denmark, Norway and Finland. Our aim is to clarify how the discursive practices reflect the current evaluation and publication policies and how the discourses construct the rationales of educational governance. We have approached our data (58 interviews) from the framework of discursive institutionalism, which sees both the underlying ideas and beliefs, and the discursive practices as the dynamic factors behind institutional change.

We argue, that in all the Nordic countries these discursive practices take place in a balancing discursive triad between global competence, neo-liberal accountability pressures and the traditions of the egalitarian Nordic comprehensive school—however with varying country-specific rationales on school accountability and transparency.

ARTICLE HISTORY

Received 17 February 2018
Accepted 11 October 2018

KEYWORDS

Educational quality assurance and evaluation; Nordic countries; accountability; transparency; discursive institutionalism

Introduction

Since the 1980s, there has been a steady growth of governing through evaluation and the use of data. The increase in the number of activities offering evidence about the performance of education in Europe and beyond is a signal of the development of *quality assurance and evaluation* (QAE) as a mode of governance within or across nation states (Henry, Lingard, Rizvi, & Taylor, 2001; Ozga, Dahler-Larsen, Segerholm, & Simola, 2011). The overall restructuring of educational systems has been drawn from the principles of decentralization, devolution and deregulation (Whitty, Power, & Dalpin, 1998). Increased autonomy at the local level, however, in many countries has meant intensified reciprocal accountability measurements at the national level (Hudson, 2007; Smith, 2016).

The question whether the range of indicators, among them national testing, should be published at the school-specific level or not, has been a heavily debated and controversial issue in many countries and also among the academic community (Allen & Burgess, 2011; Ozga, 2009; Visscher, 2001). In the Nordic countries, the national school evaluation practices institutionalized during the 1990s and the 2000s into different trajectories (Eurydice, 2009a). In

Sweden, Norway and Denmark, the governments have launched web portals which show manifold statistical indicators on a school level, while in Finland evaluation results are published only at a very general level. The differences lead one to think that there are different country-specific institutionalized ideas, rationales and discursive practices, not only on school evaluation, but also on school accountability or public information within the Nordic region.

Previous comparative country reports (e.g. Eurydice, 2009a; OECD, 2013) offer a good basis for detecting similarities and differences in countries' policy practices but 'static' comparisons (e.g. test results are published/are not published) do not easily capture the institutional processes which constantly legitimate or challenge these decisions. In this article we point out the importance of discourses, how they both reflect and construct the understanding of what is held to be reasonable, needed, useful etc. in Nordic school evaluation contexts.

Thus, by comparing the discourses of the educational experts in four Nordic countries (Sweden, Denmark, Norway and Finland, in total 58 interviews) we aim to clarify how the discursive practices reflect the current evaluation and publication policies

and how the discourses construct the rationales of educational governance. We have approached our empirical data from the framework of discursive institutionalism (Schmidt, 2008) and problematize the experts' views using two central concepts that label recent neo-liberal educational governance, namely accountability and growing demands for public information, transparency.

National evaluation policies and the use of results in four Nordic countries

In the Nordic countries, the foundations for the current QAE policies were formulated gradually during the 1990s and 2000s. Related to the 'neo-liberal turn' (e.g. Blossing, Imsen & Moos 2014), evaluation became a central mode of educational governance. The main features of the neo-liberal turn, decentralization, new public management and the introduction of market-logic actions such as school choice and privatization, have been described widely in the academic literature (e.g. Hudson, 2007; Lundahl, 2002; Ozga et al., 2011).

A brief overview shows the variety of the evaluation practices. Sweden was first to introduce a goal-oriented evaluation model with compulsory standardized testing for all pupils in 1994. The main aim was to enhance the comparability and fairness in pupils' grading and to achieve systematic information of the decentralized school system by continuous evaluation (Eurydice, 2009b). Sweden was also the first to publish school-specific evaluation results. Since 2001, the results of national testing have been published on the National Agency for Education's SIRIS internet database.¹ Defining the audience as widely as possible was a clear principle from the start:

The key social function of schools means that citizens have a democratic right to have access to this information. Child care and education affect almost everyone. In the Agency's view, public access must therefore be as extensive as possible. (National Agency for Education, 2018.)

Following Sweden in the 2000s, compulsory standardized tests for all pupils were also implemented in Norway and Denmark (Eurydice, 2009a; OECD, 2011a, 2011b). However, the decision about how to publish the results was not as straightforward as in Sweden. In Denmark, the Folkeskole Act in 2006 increased elements of evaluation and accountability in the Danish school system by requiring various kinds of documentation on schools and students' individual learning progress and making it mandatory to publish results of the school-leaving examinations of compulsory education for schools and municipalities (OECD, 2011b, p. 28). Finally, in 2016 the Ministry of Education launched a web portal

which enables the comparison of the schools across several indicators.²

Also in Norway, the emergence of a national testing system was ulterior and more delicate. The question of publication policy went through a heavy political debate, initially leaving an undecided outcome, one in which the results for individual schools in the national tests were not made available by the responsible agency itself, but by the media, since the agency was required by law to provide the information to the press on request. A few years later, the national testing results were added in the official web portal 'Skoleporten',³ on which several indicators can be observed at the national, local and school levels (Norwegian Directorate for Education and Training, 2011; pp. 16–18; Hovdhaugen, Vibe, & Seland, 2017).

Finally, in contrast to the other Nordic countries, the Finnish national testing policy and its practices institutionalized strikingly different. Despite many similar policy processes in the 1990s (decentralization, school choice), it was decided that the Finnish national testing was to be sample-based, covering roughly 5 to 15 per cent of the age cohort. A central argument for this even internationally rare practice (see Eurydice, 2009a, p. 27) was to avoid public school rankings, which were perceived as being detrimental and likely to accelerate school segregation (Jakku-Sihvonen, 2013). It is stated in the Basic Education Act (628/1998) that only 'the salient results shall be published', meaning in practice that results are published at the regional level or according to certain population factors.

The differences in the official policy guidelines on how school performance is evaluated at the national level, and especially for whom and for what uses the results are aimed at indicate that there are varying country-specific ideas, rationales and discourses that underlie these policies. In order to comprehend this diversity, we will next discuss two essential concepts, *accountability* and *transparency*.

Accountability and transparency in the governance of education

Accountability has become a cornerstone of public sector reforms in many countries. In general, the underlying rationale in accountability is that the producers are held accountable for the outcomes they generate. In the educational context, teachers and schools—who are trusted with the imperative task of teaching and instructing children—are the 'producers', while pupils' test results function as the measurable 'outcomes' (Rosenkvist, 2010). Different country-specific emphases in respect of accountability have led to defining countries as representing either 'hard' or 'soft' or 'high' and 'moderate' QAE techniques (OECD, 2013).

Features of 'hard' QAE countries consist typically of external evaluation measures, inspections, standardized high-stake testing and ranking lists, commonly combined with different outcome-related incentives, e.g. sanctions or rewards. Emphasis on self-evaluation, sample-based or thematic evaluations and the absence of school rankings feature in 'softer' QAE policies (OECD, 2013; also Ozga et al., 2011; Smith, 2016).

The concept of accountability is tied to its counterpart, autonomy. In the Nordic context, the autonomy of the local actors, municipalities and schools, was increased considerably in the 1990s decentralization process. Decentralization and autonomy, however, are usually approached as a trade-off that includes reciprocal accountability measures:

In a context of school autonomy, greater policy attention is given to areas such as school leadership, capacity for schools to self-manage (including self-evaluation and the monitoring of the quality of teaching and learning) and ability to implement improvement processes. In addition, the greater responsibilities assumed by schools imply greater accountability requirements such as external school evaluation and public reporting of student performance. (OECD, 2013, p. 45.)

Thus, a central question which follows, is *accountable for whom*, and consecutively *for what reason, why?* In other words, which audiences or stakeholders are seen as being entitled to have access to evaluation results and how is this decision legitimized in society?

To conceptualize the question *for whom*, we utilize the idea of accountability as a relational and hierarchical web (see Bracci, 2009) composed of different layers with specific stakeholders. In this view, the accountability relationships originate from below always to a higher level. The test results are produced in a class led by a teacher and in a school run by the principal. In a decentralized system the evaluation results may be used by the local authorities but because of the idea of reciprocity between autonomy (local) and accountability (state), the national level actors (ministry and policymakers) may be seen as the primary users of the test results. However, the highest level of the hierarchy can be thought of as representing society, as in a democracy, the elected political representatives are ultimately in a *political accountability* relationship with the members of society for their decisions and governance practices (Darling-Hammond, 2004).

Furthermore, a more direct form of accountability supports the idea of society as the primary results user. Here, we emphasize the concept of *transparency* as a growing ideology that affects the accountability relationships shown above. As Tero Erkkilä (2012) argues, the principle of public information has always been a core feature of the Nordic states' governance,

but during the last few decades, the connotation of public sector information has shifted globally, including in the Nordic countries, to emphasize public sector efficiency and *market accountability*.

In the educational sector, this trend has become visible through several standardized measurements and the production of public data and rankings, aimed not only at the decision makers, but more directly at the pupils' parents. At least three distinctive, yet overlapping accountability functions or rationales can be identified at the basic education level: *accountability to taxpayers*—a right of those paying for services to know the results; *accountability to parents* as school choosers—in order to make a rational school choice, the parents must be entitled to have access to relevant and comparable information on the quality of the service; and *accountability to citizens*—a more general governance principle, which highlights the undisputed democratic right of the citizens to have access governance information.

The interwoven concepts of accountability and transparency seem to be the dynamic factors which keep pushing governments towards intensified evaluation and to publish results more openly. As presented above, these four Nordic countries have all approached the questions of how to collect school performance data optimally and the extent to which the results ought to be published, in different manners, based on their own historical, cultural and political trajectories. The official legitimations were described briefly earlier, but if we want to turn our focus to more recent policies or anticipate future paths, we have to look at the discursive practices which constantly interpret, renew or challenge the legitimacy of the current evaluation practices in the society.

From ideas to discourses—discursive institutionalism (DI) as a theoretical framework

As the theoretical framework to analyse our empirical data, we have followed the idea of Vivien Schmidt (2008) on *discursive institutionalism* (DI). Whereas the other neo-institutionalist 'schools' (see Hall & Taylor, 1996) have been more prominent in explaining the continuity of institutions, Schmidt (2008, 2010)) sees the ideas and discourses as the fundamental and dynamic factors behind institutional change.

In this article, we understand the national testing practices and the publication policy as institutional forms representing the QAE culture in each society. We accept the notion of sociological institutionalism that in order to sustain their legitimacy, the institutions have to deal with a certain type of 'logic of

appropriateness' (March & Olsen, 1989), connected to the cultural norms and values in society. However, new ideas and discourses constantly shape the balance of the legitimacy.

In DI, ideas are seen as the substantive content of discourse. The ideas occur at three levels of generality: 1) as *policy solutions* providing suitable means for solving a specific problem or achieving the objectives set; 2) at a more general policy level as *programmatic beliefs* (Berman, 1998), which are more like paradigms or underlying principles to define for example the problems to be solved, the goals to be achieved and the norms, methods, and instruments to be applied; and 3) a *philosophical* level which works as the fundamental core, often remaining implicit or taken for granted (Schmidt, 2008).

The ideas become conveyed, challenged and exchanged in the discursive interaction processes between different actors and audiences. The *coordinative* discourse consists of actors at the centre of the political decision-making (politicians, civil servants, experts etc.), whereas the *communicative* discourse takes place in a wider political sphere through mass media and nowadays more through social media (Schmidt, 2008).

Despite the fact that discourses can originate from and be modified from the civil society to the administrative level, the direction of discursive interaction is often top down. Political elites tend to interweave the coordinative and the communicative discourses into a master discourse which sets out not only the visions for policymaking—what is, and what ought to be—but also defines the terms and the frames for the public discussion—what is rationalist, appropriate, how and why (Schmidt, 2008).

Aim, method and data

Thus, by comparing the coordinative discourses of the educational experts in each country, we aim to find new clarification in the discursive practices and in the legitimization processes to explain the similarities and the differences in the current comprehensive school QAE practices in the Nordic region. Our research questions are as follows: 1) how is the recent national testing policy and especially the use of the results legitimated/challenged by the educational experts in the Nordic region? 2) which underlying ideas or programmatic beliefs seem to guide these discursive practices? A special attention has been given to the Finnish case in contrast with the other Nordic countries. We assume that both the ideas and thereby also the discourses in the Finnish data differ from the other Nordic countries in respect of QAE-related rationales but also how the question of transparency is approached.

Our data comprise 58 interviews with well-versed experts in the educational field (senior politicians, civil servants, academics etc.) in each country: 17 from Sweden, 14 from Denmark, 18 from Norway and nine from Finland. The interviewees were identified through a 'snowball sample' method (also 'chain sample'), in which the informants were asked to suggest other relevant informants (Noy, 2008). The method was found useful to map the central actors in an international and comparative research setting. To begin the discussions, one main question was addressed to the informants, 'what is the story of your country's basic education?' The open interview structure provided freedom to the informants to bring up the elements of the story they found most important, evaluation being indisputably one central topic discussed throughout the data when reaching more recent events.

After a careful reading of the interview transcriptions, three discursive approaches emerged in each country's data, but with important country-specific emphases: firstly, national evaluation in relation to global competence; secondly, a (neo)-liberal pressure to increase market accountability of the education system; and thirdly, national QAE for serving the traditions of the Nordic egalitarian school systems.

We will next move to the country analyses and summarize our findings in the following chapter. In order to protect the privacy of the interviewees, all the quotations are presented anonymously.

Sweden—politicized quality and strict accountability

Overall, the question of national assessment was not a particularly contested issue in the Swedish interviews—the need for evaluation was rarely questioned nor much defended in our data. The methods and functionality of the current national QAE system were discussed in practical terms, compared with the other Nordic countries, and the question was more *how* evaluation measures should be implemented and what their consequences might be, than *why*.

In the PISA aftermath, the role of faring in international competition became more prominent in the QAE discussion, and in the Swedish data, even when talking about national evaluation, PISA tended to come up. Swedish pupils did quite well in the first PISA study, but the results steadily declined in the subsequent evaluations until more recent years. Having previously perceived themselves as 'the big country in the Nordic region' (SWE12), the declining results opened the system up to more scrutiny, and finally in 2008 an independent body of school inspections was established. The quality of education both in terms of national performance and international competition became a political issue, and as one

interviewee described it, education became a more important issue in the 2006 general elections than ever before. In the narratives, the perceived decline in the school results was not only a concern about providing a quality education for new generations, but also about poor performance in comparison to other countries, which, again, strengthens the need for national performance measures. Simultaneously, there were several critical notions about PISA, questioning whether it actually benefits national educational policies.

[I]f you look at the grades in Sweden, grade point average has gone up all the time. In the national tests the results are more or less the same each year. So, we have not seen anything new, there is no national investigation that would show that Sweden is doing worse, it is only in the international measurements. (SWE13)

The rationale behind the QAE system is that it will safeguard equality of opportunities, an idea which is closely tied to the welfare-state model's educational equality. Since pupils' grades affect advancement to further education, they should be comparable and not depend on their teacher, school, or classmates' results. Moreover, supervising the quality of schools is also central in ensuring that each student receives instruction of equal quality.

The market-driven system had a distinctive role compared to the other Nordic countries in terms of steering quality. On the one hand, the quasi-market system was based on the idea that only good schools will survive in the market, and the responsibility for monitoring the quality of schools was delegated to the customers, i.e. families. On the other hand, the presence of private actors in the market created a new need for monitoring schools, to ensure all schools, independent or public, meet the national standards—the need to re-centralize. In both aspects, transparency works as a tool for accountability: publishing school-specific results is an attempt to keep them accountable—not only to the state but also towards parents, who are then thought of as making rational and fully informed decisions on the education of their offspring. In practical terms, schools and teachers were also made accountable to parents via an appeal system, through which an unsatisfied parent-customer can report their concerns to the school authorities. Parents' increasing role was not seen as being unproblematic in the discussions, yet publication of school-specific results was not commonly considered to be a 'hot political topic', indicating that it is already a normal practice, not worth questioning.

And parents moving in [the school], which I spoke about before, for which the door was opened in the mid-90s, it has led to parents having a possibility to report. Anything that doesn't seem to work well one

can report to the Schools Inspectorate. So, schools work quite a lot today answering the Inspectorate about reports they have had. (SWE13)

The Swedish QAE discourse had a clear vein of concern: questions about having too much control and providing too few tools for development were raised. According to many informants, the need to have the basic education system and its quality under control has led to keeping teachers and schools under a microscope. In the most critical notions, the current accountability system was compared to 'a Soviet-type control system, only implemented with market-based mechanisms' (SWE06). Overall, there was a lot of concern about teachers and the fact that they are supervised too much, which in turn, has affected the popularity of the teaching profession—which, again, is a threat to the quality of teaching. Inspecting teachers' work diminishes trust and, as one informant put it, takes away their professional ownership (SWE05). Some of the informants worried that too many reforms had been undertaken in the wake of the declining PISA evaluations in the 2000s, and the focus of developing the education system now lies too heavily on test results. Another question was, if evaluation data is used too much to control and too little to develop, and whether schools and teachers get enough tools for improving their work, rather than focusing on meeting national standards.

As we got all these deregulation, decentralisation, marketisation, the state had to take new responsibilities. Be much more of a control institution or apparatus. And so today Swedish teachers they, there are more grades, more tests, more inspections. [...] There were maybe too many reforms and too much documentation because it's also about having individual plans and documentation, teachers have to sit down and write novels almost. [...] So.. the teacher-, teaching profession has gone down in popularity immensely over the years. (SWE03)

Denmark—gradual steps towards holistic accountability

Similar to in other countries, the Danish interviews were framed heavily by international comparisons, mainly the PISA assessment. Denmark was described as underachieving in the first PISA and eventually this 'shock', due to its high attention value, fostered the political imperative to reform the Danish school system by introducing a goal-oriented model with common goals and by strengthening the national quality assurance and evaluation system:

Over time, we stopped discussing the relevance of PISA measurement as a philosophical and methodological enterprise. And instead we discussed, what needs to be done with the Danish school. So in a sense, the power of PISA, regardless of all the

methodological qualms, the political and pragmatic side of PISA and its implication, so to speak, won. Because every time there is a PISA discussion in Denmark, the discussion is, 'how should we reform the school next'. It is no longer 'what is wrong with PISA'. (DEN04)

A major change in the Danish evaluative culture was mentioned as having taken place during the 2000s as the focus of the evaluation shifted gradually from the diagnostic and formative pupil assessment to emphasize school level evaluation and accountability. The OECD was seen as the 'main driver' behind the quality reforms but the province of Ontario in Canada was also mentioned as 'a heavy source for inspiration' (DEN03). The incongruence of high costs and low results was to be tackled by a more systematic and standardized evaluation despite the criticism by the teachers and their union:

So we had a visit from the OECD and they said you need to do something about the way you follow up on your students and how you evaluate. You need to have a better evaluation system.... And we came with our report in June 2006 saying that we want to have national tests, we want to have quality reports. We want it to stand in the aim that folkeskole should prepare for further education. It's not enough in itself because we are in a globalised country. And each municipality should prepare quality reports about some statistics of how schools are doing. And the Teachers Union hated this. They really hated it. (DEN05)

Fostered market accountability, publication policy and especially the quite recently launched web portal were presented in the interviews as highly debated and controversial issues. The discussions included doubts about its unintended consequences, but at the same time, publishing school results and quality reports were described as a factual—even if very limited—source of information, to which the citizens, that is parents, are entitled either in their role as a school chooser or as a controller of the flow of tax money:

For [the government] it's, they think, that by giving this access to a quite simple web-page where you can compare schools in your district and things like that, it helps parents to do school choice, and benefit from the idea of free choice of schools. That's the idea. [...] And some think that it's misleading. I don't think it's misleading. Of course it's, grades are grades. Test results are test results. So they are of course an information, I think. [...] But, it can be misleading if you only look to those. (DEN03)

Methodological issues were also raised when discussing the publishing policy. Value-added evaluation results which control the socio-economic factors of the pupil population at each school were promoted, be it with reserve. In general, many Danish interviewees had mixed feelings: on the one hand, there was mistrust of the competition logic and school

rankings, but on the other hand, an opportunity for better school development and better follow-up, and especially the role of public information, was approached as a fundamental and inescapable principle in a modern society:

So, I'm very sceptical about rankings. But I'm also extremely sceptical about the view that rankings should be prohibited. Because in a knowledge society, people who have a little bit of skills can publish whatever they like. It is no longer the idea that we have this one government that controls all the information. So, I would rather have that we undermine these rankings because we made them ridiculous, and we discuss them, and we compare them over the years. I would rather go in that direction, than having this idea that said 'you know, the people don't deserve to know because they don't have the competence to interpret these things' [...] I would say, publish it if you want to. Let's have the debate. And it's fairly ridiculous in the long run in a way. (DEN04)

Despite all the problematising, the near future trend was presumed to continue towards a higher level of accountability than changing the policy direction. The recent Danish QAE policy was even described as a self-piloting process, immune to domestic power relationships.

I think the next step we will see, is introduction of incentives. Our accountability system has been introduced slowly, step by step over the last ten fifteen years... And it doesn't matter, whether it is a right-wing or a left-wing oriented government, there's no big differences when we change government. (DEN03)

Norway—quality assurance for the 'child's best'

Most of the Norwegian educational experts described the 1990s as the starting point for a new and emerging assessment culture. As in many other countries, the 1990s meant a new kind of orientation to public education as a national economic investment. While previously Norwegians believed that they had 'the best education system in the world due to huge investments', now the focus was on cost-efficiency. Evaluation was seen as a tool to observe and clarify the 'paradox' of educational investments.

Yet, in the 1990s education politics, they didn't conclude by saying that you have to introduce a national test system. But they asked whether they can know about the quality if they don't have any kind of research or evaluation of some kind. So they asked how can you know, if you don't check it? (NOR01)

Findings from several domestic and international evaluations in the late 1990s, together with the Norwegian version of the 'PISA shock', legitimized extending the QAE processes towards a more systematic and comprehensive testing apparatus, to get research-based information in order to reveal the

'black box' (NOR17) of the Norwegian education system. As a result, national tests were introduced in 2004.

And then the PISA results came in 2001, which kind of pushed a lot of ongoing development forward. People tend to point towards the PISA results but I believe that, a lot of different elements that were pushed forward in the education reforms of the 2000s, like the evaluations showing a lot, a lack of disciplinary issues. [...] And people started to question what people learn in schools. What do people learn? We cannot only trust that we put a lot of money in there and we focus on the processes. We have to know something about the processes as well. (NOR06)

The public availability and use of the evaluation results was initially described, debated and found to be complex. Still, the publication policy did not show up as the main question in the Norwegian data. For instance, many of the interviewees couldn't describe the current publication policy. The confusion might be explained by the many gradual phases during the challenging implementation process of the testing framework described earlier (see Hovdhaugen et al., 2017).

However, even though parents and pupils are entitled to school choice, school-specific indicators were not considered to guide parents' choices excessively, even in the big cities. Instead of concealing the results, providers of education were seen as having emphasized the complex nature of the evaluation information—as something that cannot be used directly to draw simplified conclusion on 'good' or 'bad' school:

And we have spent a lot of time, and also the researchers to explain, that this is just, this tells very little. So if this is your only result, you should, you have to look into more. And you have to analyse this, in connection with what you know about the students from your class and from your other observations. So this is just one piece of information. And we have explained for many years that, it's no use in publishing the rankings. Because, the differences are so small. (NOR09)

Of all the Nordic countries, perhaps in Norway the QAE policy was most clearly discussed from the pupils' perspective, primarily to serve the pupils. Both the external assessments (national tests) and self-evaluations were described as having been developed in order to monitor education providers (schools & municipals) in doing their best for the child. This 'child-centric assessment' discourse, the historical roots of which date back to the early 1930s, works as an important legitimization mechanism between the early starting point of 'soft' and 'humanistic' assessment and the recently intensified quality assurance policy in Norway. All the evaluation practices are carried out to help and benefit the pupil:

Our education system is very different today than before. Because now we know more, we have developed, more research, more indicators. It's more about the content, it's more about the student learning. It's more focused on student learning. So I think the story from that time is... it's more on quality, quality assessment. How to make sure that all students really learn, and really fulfil their potential. (NOR09)

Finland—non-accountability through appropriate trust

The Finnish comprehensive school system got worldwide recognition as the first PISA results were published in 2001. Reaching the top positions among all participants was somewhat unexpected, a sort of positive PISA-shock compared to many other countries in which results were lower than anticipated. However, since 2006 the Finnish PISA results have indicated a gradual decline in the pupils' learning and more recently a growing variance between female and male pupils. These trends were widely discussed in the Finnish interviews as alarming signals for the Finnish school system. Interestingly, the conclusions drawn were not linked to any needs to increase control aspects or accountability on Finnish schools, but merely raised concern on budgeting or teachers' career training (FIN08). The present leading PISA performers from South-East Asia were described in an unfavourable tone as 'intensive production units' with 'teaching-to-the-test culture and long school hours' (FIN02), as if playing totally another game with different rules.

And now we have more these PISA assessments and the recent results indicate a slight decline. But we are still, according to these, the best in Europe. Not bad at all. Then there are these others, Shanghai, Korea and the rest. This competition league is a bit different, even if not knowing all the details. (FIN03)

The Finnish informants beheld a strong and a shared collective understanding of the mechanisms related to school system evaluation. The unintended consequences of high-stake testing practices and school accountability seemed to justify, even if implicitly, the reasoning for promoting continuity than change. Notably, no potential benefits of improving the Finnish school system through changing the present national evaluation practices were identified. This can be understood as one central *programmatic belief* (Berman, 1998) constructing the Finnish evaluation discourse, which is distinctive to other Nordic countries. On the contrary, introducing a more systematic evaluation system with standardized testing is simply not seen as being of benefit to the system.

Unlike in many Anglo-American countries, in which they have been seeking for better results through measurement and testing and sort of tighter school-level control, I don't believe it is our way. I don't

believe it will bring results. The question is more of how to build the school culture which will support the learning from the early childhood on. And now this segregation process we have is worrying. It is not of what the school does alone but also of how learning is valued in the society and I think in this we have gone backwards now. (FIN10)

A general concern about the direction of the Finnish comprehensive school in the near future was present in the interviews. However, the concern did not apply only to the basic education or the PISA results, but to a more general attitude towards learning and education in society. This is noteworthy because the education system and especially the comprehensive school system have been traditionally valued highly, as cornerstones for a small country to succeed in global competition. Changes in the societal environment could be expected to open new discourses in the QAE field. Here, the potential change towards increased evaluation or accountability is still articulated as not being needed because of a high level of trust among the central actors, namely the teacher, the principal and the local/national governance officials.

Our governance system is now based on the trust that the local actors follow the core curriculum for basic education. And if someone does not, what does it follow – nothing. At least never has, even it is a norm. And we like to think that in the municipalities they trust the principals and the teachers for doing their job, which they also have done. In the background we have of course our teacher education, not as good as advertised, but still fairly good so that the teachers are able to get their pupils far. So, we don't need this kind of testing system, which I think would only lead to teaching to the tests. One must remember that education is for life, learning is still an adventure. (FIN01)

The pressure to change some core elements of the Finnish QAE policy was recognized and outlined by some of the interviewees. The identity of a Finnish teacher as an autonomous actor and societally-respected professional was seen as a key factor in buffering the pressure. Only when discussing the fairness of the grading system was autonomy considered to be problematic but still subsidiary. Without any compulsory standardized national testing or examination, the equivalence of the grades given by the teachers was considered as a sort of blind spot in the Finnish school system.

And the core competences, let's say mathematics, are declining and soon it may be so low that I presume it is EK [Confederation of Finnish Industries] first who states that the comprehensive school system is collapsing and we must have a better clarity of our learning. And then we will have a testing unit waiting already for doing this task. And it is easy to see how it leads to a claim for having final exams like in the US. But for

now we don't have that. The Finnish teacher is so autonomous it is unthinkable. And it may have even gone too far, no one really knows what is going on inside the classroom. The Finnish principal does not listen out for the teacher because it just isn't right, because of the trust. The new curriculum is very loose, we don't have any inspection system like in the old days, no control for textbooks so you can just sit there and do your group work of what you wish. And then that we don't have any tests and control for comparability, we know now from various researches that the grades given for the same level of competence vary by schools, I think this problem will arise much louder, especially if we fail in the PISA. (FIN04)

In contrast to other Nordic countries, the publication policy was not discussed in the Finnish data in terms of either accountability or transparency. The absence was not surprising, as the issue has been rarely touched and the current policy seems to enjoy a wide common understanding among the decisionmakers as well as in society. This can be understood as another underlying truth or paradigm in the Finnish evaluation culture, where different actors, also parents are expected to follow a certain logic of appropriate behaviour (March & Olsen, 1989), not to challenge the autonomy of the schools. Moreover, excessive consumer activity was considered to be detrimental to the core of the Finnish comprehensive school system, to break 'the idea'.

I guess it was in 2009 when this one private tutoring company started. These provide tutoring for certain areas that have enough potential customers, say well-off families. And this is market-based, we cannot prevent the supply. But it will break the idea of the comprehensive school if the parents start to estimate that their child won't get enough impulses or guidance and start to pay for it, then it will crack. (FIN06)

The discursive triad of school evaluation practices in the Nordic

Based on our empirical findings, we argue that the discussion of school evaluation practices in the Nordic region is formed in the interplay of a discursive triad, presented in Figure 1. The three main discourses are related to three forms of accountability, specific to the Nordic context.

A *global competence discourse* concerns the overall quality of the basic education system in the context of economic competitiveness and international comparisons. This is mostly discussed when referring to the PISA results, linking lower rankings to the imperative increase of testing and accountability measures (political accountability). The importance of PISA is noted in each country, if in varying manners: in Denmark, Norway and especially in Sweden, the national discourse is more about reforming the system to perform better in international measures. Since it attracts such wide-

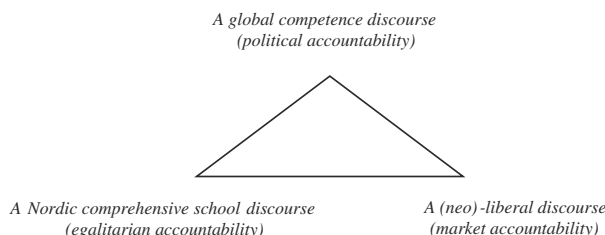


Figure 1. The discursive triad of national level QAE policy in the Nordic.

spread attention, PISA acts as a final legitimating force for national reforms. In Finland, the effects have been almost the opposite per the initial success, and despite the recent decline in the latest PISAs, the Finnish system is still seen as being competitive.

A *(neo)-liberal discourse* articulates the increased testing as a follow-up service for individual schools and pupils. The publication of school-specific evaluation results is primarily expressed as a fundamental principle of governance transparency and as subsidiary to promoting consumer behaviour and school choice (market accountability). The market accountability is strongest in Sweden, where the public nature of schools' results is built in to the quasi-market system, whereas in Finland, this discourse is most clearly absent. Overall, the Nordic countries are at different stages of the institutionalization of evaluation practices, not only structurally but also discursively. In Sweden, intensified evaluation and school-specific indicators are approached in practical terms, and in Denmark and in Norway, as a dynamic process that is still seeking its shape. In Finland, the discursive practices aim to defend the status quo from external challenges.

A *Nordic comprehensive school discourse* highlights the traditions of Nordic egalitarianism in order to prevent the increase of market-logic in education. This discourse is manifested in our data as a worry for the teachers about their status and autonomy, which external evaluation or publishing results challenges. It derives its legitimacy from the traditional idea of a common and equal Nordic comprehensive school (egalitarian accountability), working as a counterforce against the neo-liberal discourse. This discourse is strong especially in Finland and often becomes articulated by referring to warning examples of high-stakes testing.

How these three main discourses are manifested is relational to the programmatic beliefs and rationalities of school evaluation in society. For example, in Finland the whole idea of raising educational quality through intensified evaluation seems highly questionable. This rationalization has undisputedly become stronger in the 2000s because of the PISA success. Furthermore, the idea that the members of civil society construct the

highest layer in the school accountability hierarchy is not questioned or even recognized as a political issue in Finland. On the contrary, the negative effects of testing and public results are many times transmitted as taken for granted. The balance of the three discourses combined with the underlying ideas and rationales specific to each society explain why public school results in one context are seen as supporting equality and efficiency, but in another as the ultimate source of inequality.

Discussion

In this article, we have touched on a topic of growing importance, school performance and its publicity in the Nordic countries. By comparing the coordinative discourse of the educational experts we were able to detect three main discourses, which reflect but also construct the present school evaluation practices in each four country. Our analysis showed that the balance of the Finnish discourses differs substantially from the other Nordic countries, not only because of the PISA success, but also in the rationales that underlie the discursive practices on pupils testing, school accountability or transparency.

Our notions are well in line with a common understanding, Finland being an upstream case in the global trend of intensified QAE practices. By contrast, the governments in Sweden, Denmark and Norway have launched web-based data portals that include various comparable indicators on pupils' learning results, teacher-pupil ratios, school resources etc. This information is aimed at different stakeholders, not least to the parents in order to promote school comparisons and to choose a school (see e.g. www.valjaskola.se). Despite the attempts to produce a broad picture of the school environment consisting of multiple factors, school-specific data have often been compiled in the media into simplified listings and rank orders, accompanied by headlines about 'the best and the worst' performers.

The Finnish decision to carry out sample-based national testing has turned out tenable. It has served its purpose of evaluating the school system in general, but not to promote or highlight school differences, which comparisons always tend to create. If data

existed for all schools, the interest in its use comparatively would very likely grow. The only comparable pupil performance test in the Finnish education system, the matriculation exam at the end of upper secondary education, is widely reported in the media by average scores of each school every year. However, it is reasonable to assume that the demands for greater transparency, whether concerning public institutions and good governance or the private sector (e.g. tax havens, corruption etc.), will increase rather than decrease in future societies. The educational sector is not a separate part of that trajectory. Therefore, it is important to continue scrutinizing the formation of the coordinative discourse, since it sets the frames, the possibilities and the limits—what is ‘appropriate’ and how—also for the future policymaking.

The change in the governance of education through increased accountability and transparency has not yet reached Finland. According to our analysis, these pressures simply do not resonate with the Finnish core beliefs on school evaluation and its benefits. Despite the gradual decline in more recent PISA assessments, the coordinative discourse produced by the Finnish elite has been consistent in setting the boundaries for public debate. Even though the Finnish school system has been subjected to critical observation several times in the news this year, neither the guidelines for a national testing system nor publicity issues have been seriously challenged on any front. The master discourse of Nordic egalitarianism, articulated through the autonomy of the Finnish teacher, trust in the system and the detriments of high-stakes testing, accompanied by the PISA success, has so far effectively controlled the faintest neo-liberal opinions.

Notes

1. <https://sir.is.skolverket.se>.
2. <https://www.uddannelsesstatistik.dk/grundskolen/overblik>). Note: the national testing results are still ordered for restricted use only (Ministry of Education, 2016).
3. <https://skoleporten.udir.no/>.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Tommi Wallenius  <http://orcid.org/0000-0003-0956-5254>
 Sara Juvonen  <http://orcid.org/0000-0003-2798-9376>
 Petteri Hansen  <http://orcid.org/0000-0003-2209-3618>
 Janne Varjo  <http://orcid.org/0000-0003-2528-4241>

References

- Allen, R., & Burgess, S. (2011). Can school league tables help parents choose schools? *Fiscal Studies*, 32(2), 245–261.
- Berman, S. (1998). *The social democratic moment*. Cambridge: Harvard University Press.
- Blossing, U., Imsen, G., & Moos, L. (Eds.). (2014). The Nordic education model. In ‘A school for all’ encounters Neo-liberal policy. Dordrecht: Springer.
- Bracci, E. (2009). Autonomy, responsibility and accountability in the Italian school system. *Critical Perspectives on Accounting*, 20, 293–312.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106(6), 1047–1085.
- Erkkilä, T. (2012). *Government transparency: Impacts and unintended consequences*. Basingstoke: Palgrave.
- Eurydice. (2009a). *National testing of pupils in europe: Objectives, organisation and use of results*. Brussels: Education, Audiovisual and Culture Executive Agency.
- Eurydice. (2009b). *National Testing of pupils in Europe: Objectives, organisation and use of results*. Sweden, Brussels: Education, Audiovisual and Culture Executive Agency.
- Hall, P., & Taylor, R. (1996). Political science and the three new institutionalisms. *Political Studies*, 44(5), 936–957.
- Henry, M., Lingard, B., Rizvi, F., & Taylor, S. (2001). The OECD, globalisation and education policy. In *International Association of Universities: Issues in higher education*. Amsterdam: IAU Press.
- Hovdhaugen, E., Vibe, N., & Seland, I. (2017). National test results: Representation and misrepresentation. Challenges for municipal and local school administration in Norway. *Nordic Journal of Studies in Educational Policy*, 3(1), 95–105.
- Hudson, C. (2007). Governing the governance of education – The state strikes back? *European Educational Research Journal*, 6(3), 266–282.
- Jakku-Sihvonen, R. (2013). Oppimistulosten arviointijärjestelmistä ja niiden kehittämishaasteista [On evaluation systems of educational outcomes and their challenges]. In A. Räisänen (Ed.), *Oppimisen arvioinnin kontekstit ja käytännöt* (Vol. 2013, pp. 3). Opetushallitus: Koulutuksen seurantaraportit.
- Lundahl, L. (2002). Sweden: Decentralization, deregulation, quasi-markets - and then what? *Journal of Education Policy*, 17(6), 687–697.
- March, J., & Olsen, J. (1989). *Rediscovering institutions: The organizational basis of politics*. New York: Free Press.
- Ministry of Education. (2016). *Retningslinjer for brug af data fra de nationale test til statistiske og videnskabelige undersøgelser*. [Guidelines for the use of national test data in statistical and scientific research]. ISBN: 978-87-603-3075-9. Ministeriet for Børn, Undervisning og Ligestilling, 2016.
- National Agency for Education (2018). SIRIS a source of information for understanding of the world of education. Retrieved 15 Jan 2018, from, <<https://sir.is.skolverket.se/siris/?p=SIRIS:33:0>>
- Norwegian Directorate for Education and Training. (2011). OECD review on evaluation and assessment. Frameworks for Improving School Outcome. Country Background Report for Norway.
- Noy, C. (2008). Sampling knowledge: The hermeneutics of snowball sampling in qualitative research. *International Journal of Social Research Methodology*, 11(4), 327–344.
- OECD. (2011a). OECD Reviews of Evaluation and Assessment in Education: Denmark. Claire Shewbridge, Eunice Jang, Peter Matthews and Paulo Santiago.

- OECD. (2011b). OECD Reviews of Evaluation and Assessment in Education: Norway. Deborah Nusche, Lorna Earl, William Maxwell and Claire Shewbridge.
- OECD. (2013). Synergies for better learning. In *An international perspective on evaluation and assessment*. OECD Reviews of Evaluation and Assessment in Education. DOI:<http://dx.doi.org/10.1787/9789264190658-en>.
- Ozga, J. (2009). Governing education through data in England: From regulation to self-evaluation. *Journal of Education Policy*, 24(2), 149–162.
- Ozga, J., Dahler-Larsen, P., Segerholm, C., & Simola, H. (Eds.). (2011). *Fabricating quality in education*. In *Data and governance in Europe*. London: Routledge.
- Rosenkvist, M. (2010). *Using student test results for accountability and improvement: A literature review*. (OECD Education Working Papers). Paris: OECD.
- Schmidt, V. A. (2008). Discursive institutionalism: The explanatory power of ideas and discourse. *Annual Review of Political Science*, 11, 303–326.
- Schmidt, V. A. (2010). Taking ideas and discourse seriously: Explaining change through discursive institutionalism as the fourth ‘new institutionalism’. *European Political Science Review*, 2(1), 1–25.
- Smith, W. C. (Ed.). (2016). *The global testing culture: Shaping education policy, perceptions, and practice*. Oxford: Symposium books.
- Visscher, A. J. (2001). Public school performance indicators: Problems and recommendations. *Studies in Educational Evaluation*, 27(2001), 199–214.
- Whitty, G., Power, S., & Dalpin, D. (1998). *Devolution and choice in education: The school, the state and the market*. Melbourne: Acer.

